5-2018

Perceptions and Cognitive Bias in Decomposition Scoring Methods in Forensic Anthropology

Kelly Ann Sauerwein
University of Tennessee, ksauerwe@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Kelly Ann Sauerwein entitled "Perceptions and Cognitive Bias in Decomposition Scoring Methods in Forensic Anthropology." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Anthropology.

Dawnie W. Steadman, Major Professor

We have read this dissertation and recommend its acceptance:

Lowell A. Gaertner, Lee M. Jantz, Walter E. Klippel

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Perceptions and Cognitive Bias in Decomposition Scoring Methods in Forensic Anthropology

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Kelly Ann Sauerwein

May 2018

ACKNOWLEDGMENTS

While this dissertation represents my own work, there are numerous people who have contributed to its completion. First, I would like to thank my advisor, Dr. Dawnie Steadman for her consistent guidance during this process and for taking me on as a student when I was switching dissertation topics. Without her, I would not be finally accomplishing my goal of earning my Ph.D.; thank you, Dawnie. My committee, consisting of Dr. Lee Jantz, Dr. Walter Klippel, and Dr. Lowell Gaertner have been instrumental in my continued success. Lee, your advice these last 6.5 years have been invaluable and I appreciate all you have given me. Walter, I started out as your student all those years ago and you helped me to succeed. I am grateful that you continued to be on my committee and we were able to finish this journey together. Lowell, you represent both halves of my graduate career. I started this extremely crazy journey in 2005 as a doctoral student in Psychology at UC Davis and while I switched fields in 2008, I somehow still came full circle to back to social psychology. Thank you for your patience and your insight. Thank you also to Dr. Michelle Hamilton at Texas State University for starting me on this path, what seems like a lifetime ago. Thank you for your advice and guidance even after my time at Texas ended.

I wouldn’t be as sane as I was throughout this process without the help of my friends and fellow graduate students, especially Tiffany Saul (Team SauSau!), Shauna McNulty, and Alice Gooding. Thank you for talking me down and bringing the snarkiness. Graduate school is hard enough, but it would have been even harder without you. Thank you to Angela Dautartas, my friend and bjj training partner, for helping me get through the stress and frustration of grad school with bourbon, grappling, and good times. We did it! A special shout out to Jenn Renner for not only being an amazingly awesome person and fantastic chef, but for your flexibility and willingness to take care of Horatio. That may not have seemed like much, but it was a lifesaver. To my friend My-An Le: thank you for checking in on me, sending me hilarious memes, the occasional care package from 2,000 miles away, and for being my friend for the last 20+ years.

To my parents, Tom and Alana Sauerwein, thank you for your constant support. Mom, while you are not around anymore to see this dissertation completed or to see me graduate, I know that you would have been proud that I finally finished it. Thanks for always telling me that my education came first and to keep going even when it was hard. Dad, thank you for the weekly chats, rants, laughs, and the fatherly indignation when something didn’t work out the way we wanted. To my husband, Charlie, what can I say that I haven’t already? Thank you just doesn’t seem enough, but thank you; because of you I was able to succeed. And to Horatio, thank you for the cuddles, laughs, and love. You seemed to know just what I needed.

And lastly, to everyone at Gracie Barra Knoxville and Maryville Jiu-Jitsu: Thank you for the early morning rolls, the arm bars, and the friendship. Thank you for accepting me into the family and helping me to grapple away the stress and frustration that comes along with graduate school and helping me develop the patience and discipline that I sorely needed.

ABSTRACT

There is increasing recognition within the forensic science community that human examiners are prone to biases that may affect the accuracy and precision of conclusions. The analysis of decomposition to estimate the postmortem interval (PMI) is one of the important functions of forensic anthropologists, yet analytical methods that rely on the judgment of human observers, such as assessing PMI from decomposition, may be vulnerable to cognitive bias, leading to inaccurate results. No studies to date have examined the effects of cognitive bias on decomposition scoring methods. The goal of this dissertation is to understand the role of the cognitive factors of mood and motivation on the analysis of decomposition characteristics both in field and photograph contexts. Fifty undergraduate research assistants with the University of Tennessee's Forensic Anthropology Center assessed decomposition from 10 donated individuals over several months utilizing the Total Body Score method. Four graduate students who had extensive knowledge and experience with decomposition served as experts with which to compare the accuracy of observers' TBS values. Observers also completed two psychological measures aimed at assessing their motivations (Intrinsic Motivation Inventory) and mood (Positive and Negative Affect Schedule) at the time of scoring. It was expected that observers and experts would differ in both the field and photograph contexts and that mood and motivation factors would be predictive of TBS values. Hierarchical random intercept multiple regression models were conducted to assess the relationship between these measures of cognition and observers' TBS scores. Observers and experts only differed in the field context possibly due to biasing contextual information regarding the placement date of the donors. No differences were found in the photograph contexts because contextual information was not available to observers.

Observers in the field context were motivated by their perceived performance, how nervous or pressured they felt while completing the task, and the perceived difficulty of the task itself. Mood only played a small role in the photograph context where negative mood influenced scoring decisions. This study has implications for the application of TBS to longitudinal research conducted at outdoor decomposition facilities as well as to cross-sectional casework.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE:

INTRODUCTION

Forensic anthropologists have been studying decomposition characteristics for decades in the pursuit of estimating time since death in order to facilitate the identification of unknown human remains. More recently, attempts have been made to standardize these analyses in order to achieve more accurate and precise estimates (Galloway, Birkby, Jones, Henry, and Parks, 1989; Megyesi, 2001; Megyesi, Nawrocki, and Haskell, 2005; Bytheway et al., in press; Kenyhercz, Steadman, and Gundel, 2017). However, these methods still require a human examiner to determine which decomposition characteristics are critical for determining time since death. As these estimates are only as reliable as the examiners who make them, it is important to consider the potential limitations of examiners, including motivations for accuracy, pressure from peers or supervisors, and their mood at the time of examination. Understanding the cognitive state of the examiner can reveal otherwise unknown internal states that can affect a person's decision-making capabilities. Cognitions are the mental processes involved in perception, awareness, judgment, and decision-making that lead to one's understanding of the world (Dror and Fraser-Mackenzie, 2009; Rossmo, 2009). The influence of these cognitions, especially biases, on examiner performance is a current concern in the forensic sciences that has not yet been fully applied to anthropological methodology. This dissertation examines the effects of observer bias upon the accuracy of conclusions obtained from decompositional characteristics derived from the Total Body Score (TBS) system, a frequently utilized method in decomposition research (Megyesi, 2001; Megyesi et al., 2005).

Cognitive processes guide individuals' perceptions and decision-making abilities. How an individual perceives the world is based on environmental influences such as the context of the perception and the characteristics of the person or thing being perceived (Simons and Chabris, 1999; Balcetis and Dunning, 2006). Additionally, perceptions are greatly influenced by the attitudes, personality, mood, and expectations of the perceiver (Robbins and Judge, 2013). These perceptual influences play a role when judging others. Oftentimes, shortcuts are employed in decision-making in order for individuals to make faster, while still accurate, judgments. This is especially true in ambiguous situations or when information pertinent to making a judgment is limited (Dror and Fraser-Mackenzie, 2009; Coutts and Gruman, 2012; Robbins and Judge, 2013). Because it is impossible for people to incorporate every stimuli in their environment into their judgments, information that makes a target (person, object, or event) stand out is more likely to be included (Coutts and Gruman, 2012). Due to this selectivity, individuals may be vulnerable to biases.

Cognitive bias is a mental shortcut in reasoning or evaluation that can lead to inaccuracies and mistakes in decision-making (Dror and Fraser-Mackenzie, 2009; Rossmo, 2009; Robbins and Judge, 2013). There are many factors that leave individuals susceptible to cognitive biases, such as accuracy motivations, emotional pressures, and time constraints (Kunda, 1990; Dror and Fraser-Mackenzie, 2009; Rossmo, 2009; Charlton, Dror, and Fraser-Mackenzie, 2010). Confirmation bias is one of the most commonly and heavily researched cognitive biases (Jonas, Schultz-Hardt, Frey, and Thelen, 2001; Bressan and Dal Martello, 2002; Stelfox and Pease, 2005; Charmon, Gregory, Carlucci, 2009; Dror and Fraser-Mackenzie, 2009; Hart et al., 2009; Rossmo, 2009; Nakhaeizadeh, Dror, and Morgan, 2014; Nakhaeizadeh, Hanson, and Dozzi 2014). This occurs when an individual seeks out information that confirms preconceived beliefs or previous choices and they ignore or are critical of any information that contradicts those ideas or choices (Stelfox and Pease, 2005; Dror and Fraser-Mackenzie, 2009; Rossmo, 2009). Confirmation bias can also be influenced by contextual information. For example, when individuals are told that two people in a photograph are related, they will rate their physical resemblance as higher than if told they were unrelated; this effect occurs even when the people in the photo were not actually related (Bressan and Dal Martino, 2002).

Cognitive biases can occur in any situation that requires an individual to make a decision. High pressure/stressful situations, such as forensic investigations, are particularly prone to these biases since producing a correct answer or a match can have a strong influence on an individual's motivations (Kunda, 1990; Dror and Rosenthal, 2008; Charlton et al., 2010; Dror and Cole, 2010; Kassin, Dror, and Kukucka, 2013). A 2009 report by The National Academy of Sciences (NAS) concluded that forensic science disciplines that rely on human analysts could be prone to cognitive biases with results in losses of objectivity. While the goal in the forensic sciences is to aid law enforcement in the identification of perpetrators, there have also been wrongful convictions of innocent people due to faulty forensic analyses (NAS, 2009; Weigand, 2009). This highlights the danger of giving too much weight to evidence derived from imperfect or imprecise testing.

The publication of the NAS (2009) report led to an increase in studies examining cognitive bias in the forensic sciences (Charlton et al., 2010; Kerstholt et al., 2010; Page, Taylor, Blenkin, 2012; Ulery, Hickin, Buscaglia, and Roberts, 2012; Kassin, et al., 2013). The primary focus of the NAS report and subsequent studies was fingerprint (Ask and Granhog, 2005; Dror, Peron, Hind, and Charlton, 2005; Schiffer and Champod, 2007; Charlton, et al., 2010; Kassin et al., 2013) and DNA analyses (Krane et al., 2008; Dror and Hampikian, 2011; Garrett, 2011); biological anthropology, specifically forensic anthropology, was not one of the disciplines singled out. However, the field has proactively strived to increase standardization and decrease subjectivity through multiple mechanisms, including the development of scoring systems for classifying decomposition stages (Megyesi, et al., 2005; Kenyhercz et al., 2017) and ancestry estimation (Hefner and Ousley, 2014). Those few attempts at documenting bias in anthropology, especially those by Nakhaeizadeh, Hanson, et al. (2014) and Nakhaeizadeh, Dror, et al. (2014), have demonstrated that confirmation and contextual biases remain within the field, particularly associated with visual assessments of trauma, sex, age, and ancestry. Therefore, forensic anthropological analyses are not shielded from the effects of cognitive bias.

Currently there is a lack of empirical, peer-reviewed research in anthropology that examines the influence of biases on assessments. The studies that do exist tend to focus on biases within the estimation of the biological profile (Nakhaeizadeh, Dror et al., 2014; Nakhaeizadeh, Hanson, et al., 2014). However, forensic anthropologists are not solely tasked with analyzing skeletal remains. Analyzing decomposition for the purposes of estimating time since death is another concentration of biological anthropologists and as such, attempts are being made to standardize methodology. However, variation in the decomposition process exists across environments making the applicability of standard methods, training, and knowledge of the examiners in different locations difficult. There is a desperate need to understand the role biases play in altering decomposition scores because of their importance and utility in estimating time since death in forensic casework.

The main goal of this study is to determine the relationship between cognition and accuracy in anthropological decomposition scoring methods, specifically utilizing the Total Body Score system developed by Megyesi (2001). This method assesses the progression of decomposition by identifying the characteristics present throughout three body segments:

head/neck, trunk, and limbs (both arms and legs) and each segment is assigned a numerical value (i.e. score) based on the decomposition characteristics observed. The three body segment scores are summed to obtain the Total Body Score (TBS) that represents the total amount of decomposition present over the entire body (Megyesi et al., 2005). Even though this method was developed from a limited cross-section of forensic photographs, it has been applied to longitudinal field research without adapting it to the different aspects of decomposition that may not be reflected in single photographs (e.g. known scavenging, location of insect activity) or to the different influences that an examiner may be vulnerable to when they view the same body every day (Myburgh, L'abbe, Steyn, and Becker, 2013; Bytheway et al., in press). Additionally, as is the case with other forensic anthropological methods, some amount of training and experience is necessary to fully understand the decomposition process. However, it is unknown exactly how much experience is needed (Anderson, 2015; Dror, 2016). Furthermore, forensic anthropological research, like a lot of research fields, utilize students as research assistants and these students have a variable amount of experience with the specific methodology prior to their involvement. Due to this, there is a need to assess observers' vulnerability to cognitive biases and the contribution of possible cognitive processes to this error because of the ambiguity and subjective nature of decomposition assessments, as well as the variability in observers' education, training, and knowledge.

In order to measure these effects, a two-phase study was designed and carried out at the Anthropology Research Facility (ARF) at the University of Tennessee, Knoxville. The ARF provides an opportunity to observe human decomposition in a natural outdoor environment. This research explored the extent to which cognitive factors influence an observer's ability to accurately analyze decomposition utilizing the Total Body Score system by comparing observers' scores with those of experts and with their self-reported cognitive measures of motivation and mood, as measured by the Intrinsic Motivation Inventory (IMI; McAuley, Duncan, and Tammen, 1987; Ryan, Connell, and Plant, 1990; Deci, Eghrari, Patrick, and Leone, 1994) and the Positive and Negative Affect Schedule (PANAS; Watson, Clark, and Tellegan, 1998; Crawford and Henry, 2004), respectively.

RESEARCH HYPOTHESES

To date, no systematic research has been conducted that examines the role of cognitive biases in human decomposition analyses. With growing awareness of the influence of bias in the forensic sciences, it is vital that biological anthropology follow suit. With the increase in the number of outdoor decomposition facilities in the United States and abroad, research into the influence of bias in decompositional analyses needs to be more fully understood. The following hypotheses will address the goals of this study:

HYPOTHESIS 1

Observers' TBS and body segment scores will differ significantly from experts such that observers will score lower or higher on average than experts.

Given the role that experience and training may play in scoring capabilities (Anderson, 2015; Dror, 2016), it is expected that observers (mostly undergraduate anthropology students) because of their limited knowledge and experience with the TBS system specifically and the decomposition process in general, will score differently than experts, graduate students with significant experience working with human decomposition. This distinction may be shown as significant differences in scores between these groups. However, if no significant difference between experts and observers is found, this may indicate that experience and prior knowledge are not crucial to the accurate estimation of decomposition using TBS system. Furthermore, lack of difference between experts and observers would also indicate that there is insufficient evidence that observers' assessments were inaccurate and hence, either there are no observer biases or biases cannot be ascertained with the current study design.

HYPOTHESIS 2

The cognitive measures of positive and/or negative mood and motivations will be predictive of TBS values.

This hypothesis directly examines whether feelings of positive or negative mood and/or observers' ratings of their motivations to enjoy and perform well on the scoring task will impact TBS values. The hypothesis is supported if any of the cognitive variables in a hierarchical random intercept regression model are significantly predictive. If no predictive model is found, then TBS is not influenced by these cognitive variables, those measures did not adequately capture the relationship between TBS and cognition, or different cognitive variables are more influential in TBS assessment than the ones tested in this dissertation.

HYPOTHESIS 3

Observers and experts will differ in their TBS assessments of cross-sectional photographs. Additionally, the cognitive variables of mood and motivation will explain observers' TBS decisions.

Due to the fact that photographs capture a cross-section of information and therefore, do not typically portray the complete story of the decomposition process from all possible angles, as is available in the field, it is predicted that observers will score differently from experts because of less experience and knowledge of decomposition and the TBS method. Photographs present a different challenge to observers in that they provide less information about characteristics than an in-person assessment where they can examine the body from all angles. Therefore, observers may have to choose scores based on less information. However, the photo assessment may be less stressful on observers because they are completing it in a comfortable, air-conditioned room away from the weather and environment of the research facility. These challenges observers experience in scoring photographs will also be shown as different cognitive factors being significant predictors in the photos as compared to the field assessment.

RESEARCH DESIGN

This study was conducted in two phases. During Phase I, 41 University of Tennessee observers scored the decomposition characteristics of 10 donated individuals at the ARF using the Total Body Score system developed by Megyesi et al. (2005). These observers also completed measures of motivation (the Intrinsic Motivation Inventory) and mood (Positive and Negative Affect Schedule) to quantify those cognitive states. Phase I was conducted in two trials over the fall (Trial 1) and spring (Trial 2) semesters from 15 October 2016 – 30 April 2017.

Phase II of this study compared these observers' ability to accurately analyze decomposition from photographs to the in-person assessments conducted in Phase I. For this phase, 36 observers were presented with 20 photographs of donors from the ARF in various stages of decomposition. Fourteen of these photographs were full body images, while the remaining six depicted isolated body segments such as the head/neck, arms or legs only, or just the trunk. Observers were asked to provide scores for those photos that they believed contained enough information to make a scoring decision. Twenty-four observers participated in both Phases I and II while the remaining 12 only scored in Phase II and had limited experience with human remains. All scoring materials in this phase were the same as in Phase I. As previously discussed in this chapter, the TBS method was developed using photographs of bodies in various stages of decomposition. As forensic anthropologists assess decomposition in person, both in cross-sectional (e.g. casework) and longitudinal research studies, a comparison of these two scenarios, field and photos, is required.

SIGNIFICANCE OF RESEARCH

The publication of the NAS report in 2009 increased researcher's awareness of the role of cognitive bias in the forensic sciences. Since then, efforts have been made to document and combat the influence of biases. This study represents the first of its kind to understand how cognitions influence the decision-making process in decomposition analyses in anthropology. Anthropologists and other forensic practitioners who are asked to analyze decomposition must understand the influence of bias on this type of examination so that they take precautions and adopt strategies to reduce or eliminate biases. The current research model, while attempting to approximate the actual conditions in which forensic information is collected, nevertheless does not directly reflect it. This project used samples from undergraduate anthropology students on known individuals allowing for control of certain variables that is not possible in actual forensic conditions. This then provides the opportunity to better understand real-life cases of bias when they occur.

FORMAT OF DISSERTATION

This project examines the effects of bias upon the accuracy of decompositional analysis. Chapter Two provides a summary of relevant background information on cognition, detailing the decision-making process, common cognitive biases, and the effect of those biases on the forensic science community. Chapter Three provides background information of the processes associated with human decomposition, common influential factors that affect the process, and ways to measure it. Chapter Four is a detailed description of the experimental design and analytical methods of both phases of this study and Chapter Five reports the results of Phase I and Phase II. Chapter Six provides a discussion of these results and their implications for overcoming bias in forensic anthropology. Chapter Seven concludes with final remarks and suggested considerations for practitioners as well as future research.

CHAPTER TWO:

COGNITION

Cognitive science is the study of how the human mind processes and interprets information (Anderson, 2015; David, Midea, and Opre, 2004). Cognition encompasses those mental processes through which external or internal information is converted, reduced, expanded, stored, retrieved, and used (Neisser, 1976; Brandimonte, Bruno, and Collina, 2006). As a result, it involves a variety of functions, such as perception, memory, decision-making, concept formation, and problem solving that mostly occur outside of conscious awareness (Heuer, 1999; Kahneman, 2003a,b). The goal of this chapter is to introduce the themes of perception and decision-making in the cognitive sciences and describe the development of common cognitive biases. This is followed by a discussion of the applications of cognition in the forensic sciences with special attention paid to the effects of biases in the decision-making process.

PERCEPTION

Perception is the process by which individuals select, organize, interpret, and assign meaning to sensory information in their environment (Rossmo, 2009). Perception is important because much of behavior is based on individuals' *perception* of what reality is, not on reality itself (Robbins and Judge, 2013). The world as individuals perceive it is the world that is behaviorally important because people presume that the world they are consciously aware of is an accurate representation of their environment, and they act on that presumption (Simons and Chabris, 1999; Rossmo, 2009). However, what an individual perceives can be significantly different from objective reality. For example, in studies of perceptual blindness, undergraduate participants were asked to note how many times a basketball was passed between people in a video (Simons and Chabris, 1999). While they were mostly accurate in providing the number of passes, 40% of participants failed to notice a person in a gorilla suit wander into the scene, pound its chest, and walk away (Simons and Chabris, 1999; Balcetis and Dunning, 2006). People are not aware of everything going on around them because perception is biased, selective, malleable, and is influenced by psychological states and the environment (Balcetis and Dunning, 2006).

Factors That Influence Perception

How is it possible that individuals may look at the same thing yet perceive it differently? There are many factors that can influence, mold, and distort perceptions, including the specific characteristics of the perceiver (the one doing the perceiving), the target (the person or thing being perceived), and/or the environment or context in which the perception is made (Robbins and Judge, 2013). Therefore, perceptions depend not only on the physical information, but also on its relationship to the situation and specific qualities within the individual such as expectations, motives, personality, attitudes, interests, and past experiences (Balcetis and Dunning, 2006; Robbins and Judge, 2013).

Characteristics of the target, the person or thing being perceived, can also affect people's perceptions. For example, boisterous people are more likely to be noticed in a group than quiet ones (Balcetis and Dunning, 2006). Because targets are not perceived in isolation, the relationship of a target to its environment also influences perceptions, as does the tendency to associate similar things together (Robbins and Judge, 2013). For example, women, cultural groups, or members of any other group that have similar characteristics, are often perceived as being alike in unrelated ways, such as political viewpoints, intelligence, or personality characteristics. Importantly though, context also matters. The time of day at which an object or event is perceived influences our attention, as does location, or any number of situational factors. Even prior exposure to seemingly unrelated information can bias how people perceive and interpret a situation (Aarts and Dijksterhuis, 2002; Balcetis and Dunning, 2006). Aarts and Dijksterhuis (2002) had participants think about fast animals like cheetahs or slower animals like turtles prior to estimating the speed of a person walking. Those who thought about fast animals estimated the walker to be going significantly faster than those who thought about slow animals (Aarts and Dijksterhuis, 2002). In another example, a person dressed in formal attire might not seem out of place at an upscale restaurant, but that same person dressed the same way, in an introductory anthropology class, might certainly seem out of place. Neither the perceiver nor the target has changed, but the situation is different (Balcetis and Dunning, 2006; Robbins and Judge, 2013).

Making Judgments from Perceptions

Research into the application of perception has shown that shortcuts are often used to judge others, and people form perceptions about each other based on perceived beliefs, motives, or intentions (Robbins and Judge, 2013). Attribution theory tries to explain the reasons why people judge others differently depending on the meaning or source attributed to behavior. When individuals observe behaviors, they attempt to determine if it was internally (under the target's control) or externally (outside causes force the individual to behave in a certain way) caused (Kelley, 1972; Fiske and Taylor, 1991; Martinko, Harvey, and Dasborough, 2011; Robbins and Judge, 2013). For example, if someone gets a flat tire, they might attribute it to poor road conditions, like a pothole (external attribution), rather than their poor driving skills (internal attribution). That determination depends on several characteristics including distinctiveness, consensus, and consistency (Kelley, 1967; Kelley, 1972).

Distinctiveness is whether an individual exhibits different behaviors in different situations and is a way to measure if the behavior displayed is normal or unusual for that individual (Kelley, 1967; Robbins and Langton, 2001; Robbins and Judge, 2013). If it is unusual, the behavior is likely to be attributed to external causes; if not, it is a product of internal attributions (Robbins and Judge, 2013). For example, if someone only drinks alcohol when they are out with friends, their drinking behavior is high in distinctiveness. However, if they drink at any time or place, distinctiveness is low. Consensus refers to the behavior of others in similar situations; if everyone responds the same way in a similar situation, then the behavior shows consensus and implies external causation. On the other hand, if the behavior is unique to the individual, that is, everyone acts differently in a similar situation, then the behavior will most likely be attributed to internal causes (Kelley, 1967). If the drinker from the previous example is the only one of her friends to drink, then the behavior is low in consensus, but if her other friends drink, then consensus is high. Lastly, consistency of actions is examined to see if the same individual reacts the same way over time; the more consistent the behavior, the more likely it is to be attributed to internal versus external causes (Kelley, 1967). If the person only drinks when she is out with friends, her behavior is highly consistent, but if she only drinks on one special occasion, a wedding, for example, consistency is low.

These shortcuts used in making judgments can be very valuable. They allow for the rapid formation of accurate perceptions and provide valid information for making predictions (Robbins and Judge, 2013). However, just because these shortcuts can provide accurate perceptions does not mean they are foolproof. In fact, these cognitive shortcuts, such as selective perception and the halo effect, can produce significant misinterpretations.

Selective Perception

Because it is impossible for people to process and integrate everything in their environment, only certain information, or stimuli, are considered so as to reduce the amount of information to process (Coutts and Gruman, 2012). Any information that causes anyone or anything to stand out increases the probability that it will be observed and incorporated. The brain chooses which features are important based on an individual's interests, experiences, attitudes and background (Coutts and Gruman, 2012). While this allows for the rapid formation of perceptions, it runs the risk of forming an inaccurate impression; people may only see what they want to see rather than what might actually be there. This can lead to incorrect conclusions, especially in ambiguous situations.

Halo Effect. The halo effect operates when a single feature is the basis for the formation of a general impression. The feature can be anything from appearance and intelligence to religious or political affiliation (Nisbett and Wilson, 1977). For example, Asch (1946) showed that a single negative trait amongst a list of positive ones could change people's impressions of an individual. Participants in this study were given a list of traits such as 'industrious,' 'skillful,' 'determined,' 'warm,' 'intelligent,' and 'practical,' and asked to evaluate a person to whom these traits applied (Asch, 1946). They judged the person in overwhelmingly positive terms (e.g., wise, humorous, popular). However, when one trait on the list was changed from 'warm' to 'cold,' participants rated the person significantly more negatively. Asch's (1946) classic study demonstrated that a single trait greatly influenced the overall impression of the person being evaluated. These perceptual shortcuts, while mostly applied to individuals, can also be applied to objects and situations. For example having a positive experience test-driving a car can lead the buyer to believe that the manufacturer's other models are equally good (Arnold et al., 2005).

DECISION MAKING

Decision-making is a necessary component to everyday life and individuals make hundreds of decisions a day, many of them without conscious awareness (Soon, Brass, Heinze, and Haynes, 2008). The way a decision is made and the quality of the choices available is largely influenced by perception as problems must be able to be recognized and the choices must be selected and evaluated (Stewart, 2002; Robbins and Judge, 2013). Decisions are made in response to a perceived problem. In other words, a disparity exists between the current state and a desired state, requiring people to alleviate the discrepancy (Robbins and Judge, 2013). However, in order to make a decision, the problem has to be perceived; what is one person's problem may be a satisfactory state for another. To successfully make a decision, information must be processed, interpreted, and evaluated for relevance to the perceived problem. Additionally, alternatives need to be developed and considered. Both of these are dependent on an individual's perceptions and at any point in the process, perceptual errors can occur that distort, bias, and contaminate the conclusions (Robbins and Judge, 2013).

Models of Decision Making

Rational Decision-Making Model. After a problem has been identified, how does the decision get made? It is commonly thought that a rational decision-making process is the best because it is consistent, flexible, and controllable, (Shafir and LeBoeuf, 2002; Kahneman, 2003a; Bazerman and Moore, 2008; Robbins and Judge, 2013). In this model, the problem is identified, choices are recognized and weighted, alternatives are determined and evaluated, and the best solution is then chosen from those alternatives (Bazerman and Moore, 2008; Robbins and Judge, 2013). The rational decision-making model relies on several assumptions, such as individuals having all the information with which to make their decision, that they are able to identify options without bias, and then able to choose the best option (Bazerman and Moore, 2008). Unfortunately, most decisions do not follow this structure and tend to be limited to simply finding an acceptable solution to the problem rather than the best, most optimal one, or identifying a problem symptom rather than the main problem itself (Bazerman and Moore, 2008). Oftentimes, individuals are unaware that they are making suboptimal decisions (Russo, Carlson, and Meloy, 2006).

Bounded Rationality. As mentioned earlier, people's limited information processing abilities make it difficult to incorporate and evaluate all of the information needed to make optimal decisions. The common solution to this is to reduce a complex problem to a simpler one that can be more easily understood (Kahneman, 2003a,b). However, some problems may not have optimal solutions because they are too complex to begin with. Additionally, individuals often seek solutions that are satisfactory rather than comprehensive. Because of this, individuals often operate within the confines of bounded rationality. People extract simplified and essential features from a problem without encapsulating all of the complexity (Augier, 2001; Kahneman, 2003a,b; Gigerenzer, 2008). Then, a rational decision can be made within the scope of the simplified problem (Augier, 2001). For example, once a problem has been identified, a particular individual searches for decision choices and solutions; these choices are unlikely to be exhaustive. Instead, choices are identified that are easily accessible, highly visible, and represent familiar reliable solutions (Augier, 2001; Kahneman 2003a,b; Robbins and Judge, 2013). These are then reviewed until one is identified that is adequate, that is, it meets an acceptable level of performance. The solution is the first acceptable one, rather than the most optimal (Augier, 2001; Robbins and Judge, 2013).

While this may seem a perfunctory way to make a decision, especially when compared to the rational model, simplicity may be more sensible and efficient (Gigerenzer, 2008; Shah and Oppenheimer, 2008). To use the rational model, a lot of effort has to be made to gather information on all of the options and criteria available, which costs a great deal of time and energy. If there is any ambiguity, unknowns, or time limitations, the rational model may not be any more accurate than a guess; the fast and easier option of bounded rationality may be the best option (Gigerenzer, 2008; Shah and Oppenheimer, 2008; Robbins and Judge, 2013).

Intuitive Decision-Making Model. Probably the least rational way of making a decision is the intuitive decision making model (Rossmo, 2009). This is an unconscious process that relies on associations between diverse pieces of information, tends to be fast, and is "affectively charged" (i.e., it engages the emotions; Myers, 2002; Kahneman, 2003a,b; Brown, 2007; Hicks, Cicero, Trent, Burton, and King, 2010; Kruglanski and Gigerenzer, 2011). While this is not rational, it does not mean the decision is wrong or that it will contradict a rationally derived decision. In some situations, relying on intuition can improve decision-making abilities (Brown, 2007; Kruglanski and Gigerenzer, 2011). However, it should not be relied on for all situations because it is unquantifiable, so it is difficult to know when a hunch will be right or wrong. The solution instead is to supplement intuition with evidence and good judgment when possible (Robbins and Judge, 2013).

COMMON BIASES AND ERRORS IN DECISION MAKING

Even though individuals make decisions by engaging in rational and bounded rationality techniques, systemic biases and errors still influence judgments. To minimize effort, individuals tend to rely too heavily on experiences or impulses – all characteristics of intuitive decision-making (Robbins and Judge, 2013). While these shortcuts can be helpful by allowing judgments to be made quickly, they also allow error, biases, and distortions into decisions.

Intuitive decision-making typically employs the use of heuristics (Kahneman, 2003a). Heuristics are cognitive shortcuts, or rules, that are used under conditions of uncertainty to substitute simple questions for more complex ones (Kahneman, 2003a; Rossmo, 2009). While they can be effective, heuristics can lead to cognitive biases, or errors caused by simplifying information processing, which can lead to inaccurate judgments and incorrect analyses (Dror and Fraser-Mackenzie, 2009; Rossmo, 2009).

Overconfidence Bias

Overconfidence, the overestimation of an individual's skills, knowledge, and/or abilities is a very prevalent and dangerous problem in decision-making (Plous, 1993). According to Festinger (1954), people have a desire to evaluate their own abilities and often do so in relation to the abilities of others. This often results in a more positive evaluation as people prefer to be seen as better than others, not worse (Larrick, Bunson, and Soll, 2007). For example, the 'better than average' effect (Svenson, 1981; Taylor and Brown, 1988; Goethals, Messick, and Allison, 1991; Larrick et al., 2007) demonstrated that when asked to evaluate their driving ability, 90% of drivers believed they were above average. Overall, people who tend to misperceive their abilities as being better than average are more likely to be overconfident when making judgments (Fischhoff, Slovic, and Lichtenstein, 1977; Krugar and Dunning, 1999; Larrick et al., 2007). The tendency of being too confident in one's abilities might prevent people from considering alternate solutions or making plans for future problems that might arise.

Anchoring Bias

The anchoring bias is the tendency to focus on the initial information received and fail to account for subsequent ones (Simmons, LeBoeuf, and Nelson, 2010). This occurs because individuals tend to weigh the first information they receive disproportionately (Rossmo, 2009). Information available at the time of analysis determines an individual's first estimate and when information is limited or incorrect, the starting point will be skewed, potentially reducing accuracy (Rossmo, 2009). Nakhaeizadeh, Dror et al. (2014a) found that erroneous contextual information regarding age at death provided to participants prior to their analyses significantly influenced the age estimates they provided. Those given information that a set of skeletal remains were from a younger individual (25-30 years old) provided lower age at death estimates than those told the remains were from an older individual (50-55 years; Nakhaeizadeh, Dror et al., 2014a). Similar to anchoring, the tunnel vision heuristic is a further narrowing of the focus to a limited range of alternatives and where the first identified alternative that appears good enough is selected without looking at other options (Rossmo, 2009). These heuristics are not suited for solving complex, dynamic problems that require more cognitive attention, such as those found in forensic investigations.

Confirmation Bias

As mentioned above, a reliance on heuristics for decision-making typically results in error and cognitive biases. One of the most common biases is confirmation bias. In the rational decision-making model, it is assumed that individuals gather information objectively (Kahneman, 2003a,b; Bazerman and Moore, 2008). However, this is not typically the case and information tends to be gathered selectively. Confirmation bias is a special example of selective perception: information is sought that reaffirms past choices and experiences and information that contradicts those choices is discounted (Hart et al., 2009; Jonas et al., 2001). In other words, people tend to search for and accept information that confirms preconceived notions while being critical of any challenges (Stelfox and Pease, 2005; Dror and Fraser-Mackenzie, 2009), thereby gathering information that confirms previously held beliefs, desired outcomes, expectations, and social stereotypes (Dror and Fraser-Mackenzie, 2009; Rossmo, 2009). People most often engage in confirmation bias when they believe their information is accurate and strongly believe in their opinions (Dror and Fraser-Mackenzie, 2009; Rossmo, 2009). Alternatively, individuals who have a strong desire to be accurate when making decisions are less prone to confirmation bias (Dror and Fraser-Mackenzie, 2009; Rossmo, 2009). For example, in a recent study by Nakhaeizadeh, Hanson et al. (2014b), participants were shown images of skeletal remains and asked to indicate if trauma was or was not present in each image. Those given information that the images were from recent mass grave/human rights excavations indicated more trauma was present than participants who were told the remains were from an archaeological excavation or provided no context. The expectation of trauma in the mass graves context was so strong that 69% of participants in that group rated an image of a sternal foramen, a feature known to have biological, rather than traumatic, origins, as being evidence of trauma; all participants in this study had experience in biological anthropology methods and the majority (82%) had an MS or higher degree in anthropology (Nakhaeizadeh, Hanson et al., 2014b).

Participants in another study by Bressan and Dal Martello (2002) viewed photographs of adult-child pairs and were asked to rate them on facial resemblance. When led to believe that the adult and child were related, such as a parent and child, ratings of facial similarity were higher, even when the individuals depicted in the photos were not actually related. Similarly, Charmon et al. (2009) showed that when led to believe the suspect is guilty, people perceive more similarity between a suspect and a facial composite.

There are many components that lead to confirmation bias, such as a failure to seek contrary evidence that would disprove the theory and not utilizing such evidence, refusal to consider alternative ideas or theories, and the failure to account for evidence diagnosticity (i.e. some data that appear to support one theory also support others; Stelfox and Pease, 2005; Dror and Fraser-Mackenzie, 2009; Rossmo, 2009). Being critical evaluators of information and striving for accuracy are ways to reduce susceptibility to confirmation bias (Dror and Fraser-Mackenzie, 2009; Rossmo, 2009).

Factors Affecting Biases

There are many factors that leave individuals susceptible to cognitive biases, such as accuracy motivations, emotional pressures, and time constraints (Kunda, 1990; Dror and Fraser-Mackenzie, 2009; Rossmo, 2009; Charlton et al., 2010). There is intense pressure to produce a correct answer or a match in forensic investigations and this motivation can have a strong influence on an individual's susceptibility to bias (Kunda, 1990). Charlton et al. (2010) demonstrated that experienced fingerprint examiners reported a heighted emotional state during both the search for a match and upon finding one, especially during high-profile cases. These investigators reported being highly motivated to reach a decision and close a case, which can result in the decision being reached sooner and based upon less evidence than would have been expected. This motivation produces a reliance on top-down processing, which is driven by information distinct from the actual data, including contextual information, expectations, perceived *a priori* knowledge, motivations, and emotions (Dror and Fraser-Mackenzie, 2009). It can also interfere with and corrupt perception, judgment, and decision-making. These types of processes result in rapid assessments, but increased subjectivity affects decision-making thresholds and increases the effect of erroneous initial information (Kruglanski and Freund 1983; Freund, Kruglanski, and Shpitzajzen 1985; Dror and Fraser-Mackenzie 2009).

Countermeasures Against Bias

While it is difficult to avoid all forms of bias in decision-making, Robbins and Judge (2013) provide several strategies to help reduce biases. These include focusing on goals to keep track of what information is relevant and irrelevant, actively looking for information that directly contradicts beliefs and assumptions to counteract the overconfidence and confirmation biases, and increase options and alternatives when making decisions to better the chances of finding the best one (Robbins and Judge, 2013). Employing biasing countermeasures such as accountability (i.e., holding an individual responsible for errors) may reduce this problem. However, while reducing errors, this does not eliminate biases entirely (Robbins and Judge, 2013).

Another countermeasure thought to combat bias is the use of logical reasoning. For example, forensic investigators could be made aware of biases, instructed to be open to new hypotheses, and avoid biasing influences such as extraneous information, stereotypes, or preconceived notions (Rossmo, 2009). Unfortunately, even logical reasoning is not immune to the effects. Evans, Barston, and Pollard (1983) revealed that individuals were more likely to express support for a series of statements only if they were believable, and not if they were logically valid. This is because people avoid cognitive strain, preferring the faster and less taxing process of relying on what is believable and satisfactory (Evans et al., 1983; Rossmo, 2009).

Kassin et al. (2013) targeted the two specific environments where it is critical to reduce bias and its consequences – the laboratory and crime scenes where evidence is collected and analyzed, and in court where the analyzed data is evaluated. They suggested a series of methods for reducing bias in the laboratory, such as: 1) working in a linear, rather than a circular, fashion where examiners should initially evaluate evidence and document their findings before making comparisons against any targets or suspects. In this way, any influence from the target, the person or thing being compared, can be reduced and the initial analysis of the evidence in isolation lacks the influence of a comparison to a target (Kassin et al., 2013; Dror, 2009). Kassin et al. (2013) also recommended conducting blind testing where examiners are shielded from contextual information that may be extraneous to their analyses, but may nonetheless taint their conclusions, such as eyewitness information or whether a suspect has confessed. Verification of conclusions should also follow a blind or double blind procedure, whereby the verifier does not know the original conclusion or extraneous information, is not chosen by the original examiner, and does not disclose their identity. This would reduce pressure to agree with the original examiner's decision. Lastly, Kassin et al. (2013) recommend that all forensic science programs include training in basic psychological concepts such as experimental methods, and aspects of social influence, judgment, decision-making, and perception as they apply to the forensic sciences.

COGNITION IN THE FORENSIC SCIENCES

The term 'forensic sciences' encompasses a broad range of disciplines, each with their own methods and techniques, reliability, error rates, general acceptability, and published material. Some of these disciplines are laboratory-based, such as DNA, toxicology, and drug analyses, while others are based on expert interpretation of observed patterns (bite marks, fingerprints); some require the expertise of trained scientists (chemists or biologists), others involve a mixture of both scientists and non-scientists trained in law enforcement techniques (crime scene investigators), medicine, or laboratory methods (technicians). Therefore, the term ‘forensic science’ is applied to a broad array of activities, each with their own potential to be subjected to external pressures and influences leading to bias and errors (Risinger, Saks, Thompson, and Rosenthal, 2002; Dror and Charlton, 2006; Dror, Charlton, and Peron, 2006; Krane et al., 2008; NAS, 2009; Kassin et al., 2013). In many scientific disciplines, a human examiner is the main instrument of analysis and as such, these analyses may be significantly influenced by psychological factors, such as time pressures, expectations, motivations, and contextual information (Dror and Rosenthal, 2008; Rossmo, 2009; Charlton et al., 2010; Dror and Cole, 2010; Kassin et al., 2013; Nakhaeizadeh, Dror et al., 2014; Nakhaeizadeh, Hanson et al., 2014).

National Academy of Sciences Report

In the forensic sciences, awareness of cognitive biases has steadily increased since a 2009 report by The National Academy of Sciences (NAS) concluded that forensic science disciplines that rely on human interpretation could be prone to cognitive biases and result in loss of objectivity. The NAS (2009) report also stated that problems with reliability, error, accuracy, and standardization allow for these biases. To combat this, NAS recommended the implementation of rigorous standards and protocols that would reduce subjective interpretations and ambiguity (NAS, 2009).

While much of the analysis and methodologies in the forensic sciences have aided law enforcement in the identification of perpetrators of numerous crimes, there have also been wrongful convictions of innocent people due to faulty forensic scientific analyses (NAS, 2009; Weigand, 2009). This highlights the danger of giving too much weight to evidence derived from imperfect testing and analyses or from imprecise or exaggerated expert testimony (NAS, 2009). One challenge facing the forensic science community in addressing these dangers is the presence of great disparities in funding, access to proper analytical instrumentation, the availability of well-trained personnel, certification, accreditation, and oversight among forensic science disciplines and among federal, state, and local law enforcement agencies (NAS, 2009). In other words, the forensic science disciplines and laboratories are not equal when it comes to resource availability. As a result, the reliability and quality of information arising from the examination of forensic evidences varies considerably across the country (NAS, 2009). Another challenge associated with the inequality in the forensic sciences is the lack of standardization in operational principles and procedures between and within jurisdictions; at the time of the NAS report in 2009, there was no uniformity in laboratory accreditation procedures or in the requirement for certification of forensic science practitioners. While SWG standards – scientific working groups that met to decide on the best practices for a discipline – were sometimes present, they were often vague and unenforceable. In general, the NAS concluded that the quality of forensic procedures varied due to the lack of adequate training, certification, standardization, and effective management (Giannelli, 2007; NAS, 2009).

The NAS (2009) organized forensic disciplines into categories according to the type of evidence they produced: analytical (DNA, serology, fiber, toxicology, fire and explosives analysis), digital, and pattern/experiential (fingerprints, tool and bite marks, bloodstain patterning, handwriting). It is the latter category that is the most vulnerable to misidentification and bias. Pattern or experience-based evidence typically involves matching a sample to a particular individual (i.e. the suspect) or classifying a sample into a particular category. These methods are based on the idea that characteristic markings are obtained by a source item, such as rifling or grooves in gun barrels, and that uniqueness is transferred from the source to the evidence being examined, such as a bullet (Saks and Koehler, 2005). When the evidence and its potential source are compared, a positive result indicates that the evidence originated from that source to the exclusion of all other possibilities (Saks and Koehler, 2005; NAS, 2009). With the exception of DNA, no other method has been exhaustively demonstrated to consistently display a connection between evidence and a specific source, either an individual or an object (Saks and Koehler, 2005; NAS, 2009). This is potentially due to differences within the pattern-based evidentiary disciplines with regards to interpretation. For example, fingerprint analyses are based on more established protocols and research than bite marks (Saks and Koehler, 2005). The NAS (2009) points out that the disparities between forensic disciplines underscores the fact that the interpretation of forensic evidence is not always based on scientific, peer-reviewed studies. While research has been done in some disciplines (i.e. Miller, 1989; Risinger et al., 2002; Dror and Charlton, 2006; Dror et al., 2006; Krane et al., 2008; Nakhaeizadeh, Dror et al., 2014; Nakhaeizadeh, Hanson et al., 2014), there is a noticeable lack of peer-reviewed studies establishing the validity of several forensic methods (Haber and Haber, 2008; Mnookin, 2008; NAS, 2009).

Because of these limitations, the NAS (2009) emphasized the need for forensic research to establish limits and standards of performance. The first step in their guidelines was to clarify the questions an analysis is to address. While some techniques may be too unspecific to permit actual identifications of a particular individual or object, they may still provide useful and accurate information. For example, while microscopic hair analysis may not be able to match a sample with a specific individual, it may still provide information about the individual from whom the sample originated (Saks and Koehler, 2005). While establishing the appropriate question is a first step, a solid research basis is required to establish the limits of performance and address sources of variability and bias. The NAS (2009) recommended that all forensic science fields develop stringent protocols to guide subjective interpretations as well as to develop rigorous research and evaluation programs. Since 2009, the increase in research on these impacting factors has increased dramatically (Charlton et al., 2010; Kerstholt et al., 2010; Page et al., 2012; Ulery et al., 2012; Found and Ganas, 2013; Kassin et al., 2013; Bieber, 2014; Nakhaeizadeh, Dror et al., 2014; Nakhaeizadeh, Hanson et al., 2014).

The issue of bias has implications that extend to the admission of and reliance upon forensic evidence in court. According to the NAS (2009), forensic evidence should be utilized only if: 1) the specific forensic discipline is based on consistent and valid scientific procedures that enables its practitioners to accurately analyze evidence and report findings, and 2) the amount which practitioners rely on human interpretation that may be prone to error, bias, or the lack of standard protocols is minimized. Therefore, the qualifications of the expert and the reliability of the evidence and its collection/analysis procedures are of great importance.

Amongst the NAS's (2009) 13 recommendations for improving the field of forensic sciences was the establishment of standardized terminology and reporting practices; increased research to address issues of validity, accuracy, and reliability; encourage research into the effects of biases and error in forensic examinations including, but not limited to the effects of contextual bias (e.g. influence of investigator theories, suspect information, etc.) and characterization of sources of human error; establishment of standard operating procedures to minimize potential bias and errors in forensic practice; the development of tools and protocols to advance validation, measurement, reliability, and information sharing; proficiency testing in the forensic disciplines, and developing quality control and assurance procedures to identify mistakes, frauds, or bias (NAS, 2009). Their recommendations highlighted potential areas or scenarios where human error and/or bias might influence examiner accuracy. The NAS suggestions also included ways to reduce bias, improve reliability, and counter the human factor in evidence analysis.

The publication of the NAS (2009) report led to numerous studies examining cognitive bias in the forensic sciences. Citing the NAS report, Kassin et al. (2013) reviewed research on primacy, expectancy, and observer biases in several forensic science disciplines. Fingerprint identification, for example, is problematic because there are no quantitatively precise objective measures and no instruments of measurement. Oftentimes, it is a latent print or partial print found at a crime scene that is compared with an exemplar from a specific suspect. While identical unrelated prints have yet to be found, an examiner must determine if the samples are sufficiently similar to conclude they both are from the same source (Kassin et al., 2013). Because there are few objective standards for print identification, inconsistency between and within examiners is a problem (Kassin et al., 2013). While two examiners might come to different conclusions, the same examiner analyzing the same data on different occasions, may also arrive at different conclusions (Ulery et al., 2012; Kassin et al., 2013). Ulery et al. (2012) found, in their study of 72 latent print examiners, that 10% of the time, the same examiner reached different conclusions when asked to examine prints they had previously analyzed. While 10% pales in comparison to the approximately 90% consistency rate, these false positives may still result in wrongful convictions of innocent people. Much of the variability was attributed to difficult cases, where more ambiguity existed or where the information available to reach a conclusion was inconclusive (Ulery et al., 2012). This lack of reliability demonstrates that the identification process can be subjective and susceptible to bias from other sources.

Page et al. (2012) reviewed research on the various types of contextual effects and biasing influences that impact bite mark analysis in forensic odontology. They concluded that bite mark analyses are abundant in sources of potentially biasing influences such as emotional effects from interacting personally with victims to collect bite mark evidence, the labeling of evidence as coming from the “victim” versus the “defendant,” and the shifting of judgment after repeated and lengthy exposure to a particular piece of set of evidence. For example, when an examiner gradually begins to see similarities between the bite mark and the suspect’s dentition sample after a lengthy analysis, an association is made that may not actually exist (Page et al., 2012). To minimize these sources of bias, analysts should attempt to interact as little as possible with the victims, law enforcement officials, and legal professionals and conduct their analyses independent of those sources. Separating collection and analysis responsibilities between multiple examiners would not only reduce the emotional impact on the examiners, but also limit the amount of extraneous, possibly biasing, information to which they are exposed (Page et al., 2012).

Kerstholt et al. (2010) empirically investigated the presence of confirmation bias in comparisons of ballistic evidence. Six firearms experts were presented with two bullets for comparison and either neutral or biasing case information. Each examiner compared the bullets to determine if they were fired from the same firearm. The biasing case information indicated that there was only one firearm used whereas the neutral information mentioned multiple firearms so there was less expectation that the bullet was associated with the crime. Kerstholt et al. (2010) found that there was no effect of the information, either biased or neutral, on the conclusions of the examiners. The biased description that the two bullets provided were most likely fired from the same firearm did not result in higher accuracy ratings than the implication that multiple firearms and crime scenes were involved. While they did not find an effect of information type, the authors did find large individual differences between examiners on two of the trials such that examiners' conclusions ranged from certainty that the samples were related to the possibility that they were unrelated. Kerstholt et al. (2010) attributed this to the level of experience of each examiner – more experienced examiners may have allotted less time to the experiment than their less experienced counterparts – and that the manipulation of information may have been too subtle and not strong enough to invoke cognitive bias.

While much of the NAS's attention, as well as that of subsequent studies, focused mainly on forensic disciplines such as fingerprint analyses (Ask and Granhog, 2005; Dror et al., 2005; Schiffer and Champod, 2007; Charlton et al., 2010; Kassin et al., 2013), DNA interpretations (Krane et al., 2008; Dror and Hampikian, 2011; Garrett, 2011), handwriting examinations (Found and Ganas, 2013), fire investigations (Bieber, 2014), ballistics comparisons (Kerstholt et al., 2010), and odontology (Page et al., 2012), anthropology is only mentioned five times in the report and none of those concern standards or bias in the field. However, anthropology has endeavored in recent years to increase regulation and objectivity by developing standardized scoring systems for classifying decomposition stages (Megyesi et al., 2005) as well as to reduce subjectivity in ancestry estimation (Hefner and Ousley, 2014). However, recent symposia at the American Academy of Forensic Sciences (Palmbach, Hammond, Rennison, and Ross, 2014; Simon, Thompson, Kassin, Sulner, and Scheck, 2015) as well as recent articles by Nakhaeizadeh, Dror et al. (2014) and Nakhaeizadeh, Hanson et al. (2014), have demonstrated that biases remain within anthropology, particularly associated with visual assessments of trauma, sex, age, and ancestry. Therefore, forensic anthropological analyses are not shielded from the effects of cognitive bias. Many analyses that rely on the judgment of an examiner to compare, for example, sex or age of an unknown individual against that of a standard, are prone to accuracy or confirmation biases (Nakhaeizadeh, Dror et al., 2014; Nakhaeizadeh, Hanson et al., 2014).

Psychological assessments are surveys designed to measure unobservable constructs or variables. A useful measure must be both valid (i.e. the test measures what it is supposed to measure) and reliable (i.e., consistency; the test produces similar results under similar conditions). The scales utilized in this dissertation employ a response scale known as a Likert scale, where subjects respond by selecting their level of agreement or disagreement with a set of statements on a symmetric agree-disagree format (Dawes, 2008; Heiberger and Robbins, 2014). The scales then capture the range of intensity of feelings for a particular subject. Likert scales can range from one to ten, although five or seven-point scales are most common (Dawes, 2008) and this scaling assumes distances between each point are equal:

1	2	3	4	5	6	7
not at all			somewhat			very true

1	2	3	4	5
very slightly or not at all	a little	moderately	quite a bit	extremely

Each item is a statement that the subject is asked to evaluate by giving the quantitative value corresponding to their level of agreement/disagreement with a balanced scale having an equal number of positive and negative responses (van Alphren, Halfens, Hasman, and Imbos, 1994; Dawes, 2008; Heiberger and Robbins, 2014).

For this study, motivation must be measured in terms of observable and self-reported cognitive (e.g. perception), affective (e.g. subjective experience), and behavioral (e.g. performance) responses because it is a psychological construct that cannot be observed or recorded directly (Toure-Tillery and Fishbach, 2014). Specific measures of motivation may distinguish between multiple aspects of motivation (e.g. intrinsic versus process-focused). For example, how quickly a person completes a task can have several motivation-related interpretations. Working slowly could mean that the individual has low motivation to complete the task (outcome-focused motivation); that their motivation to engage in the task is high such that they are "savoring" the task (intrinsic motivation); that their motivation to do it correctly is high such that they are highly focused (means-focused motivation); or even that they are tired (diminished psychological resources; Toure-Tillery and Fishbach, 2014).

In previous research on bias in forensic anthropology, participants were given a task such as assessing sex or the presence/absence of trauma (Nakhaeizadeh, Dror et al., 2014; Nakhaeizadeh, Hanson et al., 2014), and provided some kind of contextual information aimed at biasing their results. These studies then assessed accuracy of the results and examined if contextual information led to biased results (e.g. saying there was more trauma present when examiners believed skeletal remains were associated with a mass grave context than an archaeological context or when no context was given; Nakhaeizadeh, Dror et al., 2014). However, this only determines that bias is present and possible contextual causes for that bias. It does not examine why individuals were compelled to utilize that biased information in their analyses. No anthropological study to date has examined the motivations, perceptions, and decision-making processes of examiners as related to bias. While demonstrating that bias exists is an important first step, in order to fully combat bias, it must be better understood *why* bias occurs and what cognitive mechanisms are at play when anthropologists, or any forensic examiner, is faced with ambiguous information. Therefore, the Intrinsic Motivation Inventory (IMI; Ryan, 1982; Ryan, Mims, and Koestner, 1983; Plant and Ryan, 1985; Ryan et al., 1990; Ryan, Koestner, and Deci, 1991) and the Positive and Negative Affect Schedule (PANAS;

Watson et al., 1988) are utilized in this study in order to assess participants' motivations and moods at the moment of scoring.

Intrinsic Motivation Inventory (IMI)

The IMI is a multidimensional, self-reported measurement scale that assesses an individual's subjective experiences while performing a specific task. Originally, this measurement quantifies individuals' interest/enjoyment, perceived competence, effort, value/usefulness, feelings of pressure/tension, and perceived choice during a task, yielding six subscale scores (Table 1). However, there are several modified versions of the IMI that have shown strong validity with as few as three subscales and only nine items (McAuley et al., 1987; Ryan et al., 1990; Deci et al., 1994). The interest/enjoyment subscale is the longest because it is considered the direct self-reported measure of intrinsic motivation. The other subscales are included because they are either theorized to be positive or negative predictors of intrinsic motivation, such as the perceived choice, perceived competence, and pressure/tension scales, or are relevant to questions of motivation such as effort and value/usefulness scales (Table 1). The IMI has been shown to be psychometrically robust and stable across multiple settings, can be modified easily to fit different task activities, and is easy to administer (McAuley et al., 1987; Ryan et al., 1990).

All subscale items of the IMI have shown to be analytically stable across a variety of tasks and conditions (Ryan et al., 1983; McAuley et al., 1989; Ryan et al., 1990; Deci et al., 1994). Furthermore, all items correlate with their corresponding subscale factor (perceived choice, competence, pressure/tension, interest/enjoyment, effort/importance, value/usefulness, relatedness) at a minimum of 0.6 and do not correlate with other subscale factors at a value greater than 0.4. Additionally, there are negligible order effects of item presentation and the

Table 1: Intrinsic Motivation Inventory containing all six subscales. (R) indicates items that are reversed scored (Ryan, 1982; Ryan et al., 1990)

Subscale	Items
Interest/Enjoyment	I enjoyed doing this activity very much This activity was fun to do. I thought this was a boring activity. (R) This activity did not hold my attention at all. (R) I would describe this activity as very interesting. I thought this activity was quite enjoyable. While doing this activity, I was thinking about how much I enjoyed it.
Perceived Competence	I think I am pretty good at this activity. I think I did pretty well at this activity, compared to other students. After working at this activity for a while, I felt pretty competent. I am satisfied with my performance at this task. I was pretty skilled at this activity. This was an activity that I couldn't do very well. (R)
Effort/Importance	I put a lot of effort into this. I didn't try very hard to do well at this activity. (R) I tried very hard on this activity. It was important to me to do well at this task. I didn't put much energy into this. (R)
Pressure/Tension	I believe I had some choice about doing this activity. I felt like it was not my own choice to do this task. (R) I didn't really have a choice about doing this task. (R) I felt like I had to do this. (R) I did this activity because I had no choice. (R) I did this activity because I wanted to. I did this activity because I had to. (R)
Perceived Choice	I believe I had some choice about doing this activity. I felt like it was not my own choice to do this task. (R) I didn't really have a choice about doing this task. (R) I felt like I had to do this. (R) I did this activity because I had no choice. (R) I did this activity because I wanted to. I did this activity because I had to. (R)
Value/Usefulness	I believe this activity could be of some value to me. I think that doing this activity is useful for ______________________ I think this is important to do because it can ____________________ I would be willing to do this again because it has some value to me. I think doing this activity could help me to _____________________ I believe doing this activity could be beneficial to me. I think this is an important activity.
Relatedness	I felt really distant to this person. (R) I really doubt that this person and I would ever be friends. (R) I felt like I could really trust this person. I'd like a chance to interact with this person more often. I'd really prefer not to interact with this person in the future. (R) I don't feel like I could really trust this person. (R) It is likely that this person and I could become friends if we interacted a lot. I feel close to this person.

inclusion or exclusion of individual subscales appears to have no impact on the other scales. This has allowed for easy modification of the IMI to allow experimenters to choose only those subscales that are relevant to their constructs of interest. For the purposes of the current research, observers' subjective experiences of the scoring task were assessed across four subscale areas: interest/enjoyment, perceived competence, effort, and tension/pressure (Appendix A).

Positive and Negative Affect Schedule (PANAS)

The Positive and Negative Affect Schedule, or PANAS, is a 20-item self-report survey comprised of a list of emotion-related traits that asks participants how much they identify with each emotion at a particular time period (e.g. in the last week, month, or in the present moment), utilizing a five-point Likert-type scale ranging from one (very slightly or not at all) to five (extremely). The PANAS specifically measures mood on two dimensions: positive and negative affect. Positive affect (PA) represents how much a person feels alert, enthusiastic, and active. High PA is a state of high energy, full concentration, and pleasurable engagement, whereas low PA is characterized by sadness and lethargy (Watson et al., 1988). Negative affect (NA) is a measure of subjective unpleasant engagement that includes a variety of aversive mood states, including disgust, contempt, fear, guilt, nervousness, and anger. Low NA is typified by calmness and serenity (Watson et al., 1988). These two factors are related to individual differences in positive and negative emotionality (Watson and Clark, 1984; Tellegan, 1985); each subscale (i.e. PA and NA) consists of 10 items.

Both the PA and NA subscales of the PANAS have been demonstrated to be internally consistent, reliable, and valid (Watson et al., 1988; Crawford and Henry, 2004). While the instructions allow for modification of the timeframe of interest, the specific traits listed cannot be altered without reassessing the scale's psychometric properties. Additionally, Watson et al.'s

(1988) assessment of the PANAS noted that when short-term instructions were used (i.e. participants were asked about their mood ‘right now’ or ‘today’), the results were prone to fluctuations, while longer time periods demonstrated greater stability. This is expected and is ideally suited for the goals of this study as fluctuations in mood are hypothesized to have a role in accuracy levels in the scoring task and could contribute to bias in participants’ ratings such that high PA or low NA individuals may enjoy the task more and be more focused, whereas low PA or high NA observers may not put much effort into scoring or not be motivated to be accurate. Both the IMI and the PANAS measures are easy to administer, understand, and complete by the observers.

As previously mentioned, the NAS report did not address specifically the procedures in anthropology in need of bias monitoring. As the focus of this dissertation is decomposition, it is important to not only understand the biological process of decomposition, but the analytical techniques utilized and the potential for bias therein; the next chapter will discuss these issues.

CHAPTER THREE:

DECOMPOSITION

Taphonomy, the physical and chemical modifications that affect the body after death, has emerged as an important tool for the reconstruction of depositional context and the postmortem interval (Wilder, 1923; Efremov, 1940; Haglund and Sorg, 1997). Decomposition, or the process by which both the hard and soft tissues in the body breakdown over time, is one such taphonomic modification. Environmental, biological, and chemical factors can affect how human remains decompose and various psychological and cognitive processes may affect how it is assessed. This assessment can aid in estimating the postmortem interval (PMI), which is a fundamental component of forensic anthropological casework. An accurate PMI estimate can aid law enforcement's evaluation of suspect alibis and facilitate the identification of an unknown set of remains by narrowing the potential pool of missing individuals who can be compared for the identifications. However, if that estimate has been compromised by bias, then the possible identification may also be invalid (Nakhaeizadeh, Dror et al., 2014; Nakhaeizadeh, Hanson et al., 2014).

As discussed in Chapter 2, forensic anthropological analyses are not shielded from the effects of cognitive bias. Many analyses that rely on the judgment of an examiner to compare, for example, sex or age of an unknown individual against that of a standard, are prone to accuracy or confirmation biases (Nakhaeizadeh, Dror et al., 2014; Nakhaeizadeh, Hanson et al., 2014). Time since death estimates that utilize decomposition characteristics are no exception since there are few standard measures and ambiguity may be high because decomposition does not always progress in the sequential manner represented by the measures (e.g. Megyesi et al., 2005).

STAGES OF DECOMPOSITION

After death, the body progresses through several stages of decomposition that can be broadly categorized as early, advanced, and late stages. The early, or "fresh," stage of soft tissue decomposition represents the events that follow in the hours immediately after death and include the processes of autolysis as well as rigor mortis, livor mortis, and algor mortis (Rodriguez and Bass, 1983; Marks, Love, and Dadour, 2009). The processes that begin early after death and continue for days or weeks afterward and are some of the most identifiable characteristics of decomposition are those in the advanced, or peak, stage of decomposition and include bacterial proliferation and soft tissue breakdown, also known as putrefaction and decay, respectively (Galloway et al., 1989; Janaway, 1996). Finally, the late stage of decomposition is when skeletonization and the eventual breakdown of bone occurs (Galloway et al., 1989; Megyesi et al., 2005).

Early Decomposition

Autolysis. The events of early decomposition, known as autolysis, begin in the immediate period after death and include the appearance of rigor, livor, and algor mortis, isolated color changes, and skin slippage. Autolysis is the process of massive cell death that results in the loss of cellular integrity and widespread necrosis (Marks et al., 2009; Zhou and Byard, 2011). Due to the loss of adenosine triphosphate (ATP) and the cessation of circulation upon death, an increase in enzymatic activity of the cells occurs. The digestive enzymes within the cells that are present in life become more active and then begin to consume the cellular material around them (Marks et al., 2009; Mayer, 2012). This process typically begins in the abdominal cavity in the pancreas, liver, and stomach, which are rich in digestive enzymes (Wilson-Taylor, 2013). Externally, autolysis is characterized by skin slippage and localized color changes, such as hues of red, gray, or green (Marks et al., 2009). Skin slippage is where the epidermal layer of skin begins to separate from the dermal layer below and, with the assistance of insects, can encompass the entire body (Marks et al., 2009; Mayer, 2012). When the skin of the hands sloughs off, also known as gloving, it can be a source of possible identification from the fingerprints retrieved from the gloved skin (Wilson-Taylor, 2013). Discoloration of the skin also begins during autolysis. However, it is during the subsequent process of putrefaction that discoloration becomes widespread (Marks et al., 2009).

Autolysis is the catalyst that begins the changes seen in the mortis triad - rigor, algor, and livor mortis. Because autolysis begins deep within the body, the external signs, such as color changes or skin slippage, are not seen immediately and, in fact, are typically present after the appearance of the mortis triad. After the first 48-72 hours postmortem, time since death estimates become less accurate and less narrowly defined because the processes that follow the mortis triad are much less predictable (Marks et al., 2009).

Rigor Mortis. Rigor mortis is the postmortem stiffening of muscles due to the buildup of lactic acid (Janaway, 1996). After death, respiration ceases, causing a depletion of oxygen that is critical for the production of ATP, which is the main energy source for the body's cells. This process begins approximately two to four hours after death and is caused by the muscle fibers' inability to relax due to the depletion of ATP; relaxation does not occur until the muscles begin to decompose and the fibers loosen (Harle, 2012; Mayer, 2012; Christensen, Passalacqua, and Bartelink, 2014). While it is dependent on ambient temperature, disease, and activity level prior to death, rigor tends to be fully developed within eight to twelve hours postmortem, and then begins to recede after approximately 24 hours; after 36-72 hours, rigor is typically dissipated (Janaway, 1996; DiMiao and DiMiao, 2001; Marks et al., 2009). This process initially progresses in a downward direction starting with the muscles of the face and neck, upper limbs, trunk, and ending in the lower extremities; it typically recedes in the same order (Harle, 2012, Mayer, 2012, Christensen et al., 2014).

There are several factors that can alter the progression of rigor mortis. Ambient temperature, for example, can increase and decrease the rate of onset of rigor mortis depending on if it is hot or cold, respectively (Harle, 2012). For example, rigor dissipates more rapidly in warmer weather, but can persist longer when the body is exposed to colder temperatures (Harle, 2012). Additionally, physical exertion prior to death can accelerate the spread of rigor mortis because the lactic acid would be built up in the muscles and they would already be depleted of ATP and oxygen (Gill-King, 1997; Spitz and Fischer, 1993).

Algor Mortis. Algor mortis is the postmortem lowering of body temperature. Body temperature typically decreases about two degrees per hour in first 12 hours after death, and then decreases one degree per hour thereafter until the body equalizes to the ambient temperature (Perper, 1993). However, the rate of cooling is easily affected by temperature such that cold temperatures accelerate the rate of cooling, while higher ones lead to slower cooling (Perper, 1993; DiMaio and DiMaio, 2001). Other factors such as the presence of a high fever or intense physical activity prior to death, can cause the body to have a higher starting temperature than the 98.7°F that represents average body temperature during life (DiMaio and DiMaio, 2001). A higher starting temperature can cause any time since death estimates based off of algor mortis to be inaccurate and more difficult to calculate. Furthermore, algor mortis can also be affected by body mass. For example, an individual with more body mass (e.g. obese) will cool at a slower rate due to the insulating effect of adipose tissue, but an individual with more surface area to body mass (e.g. emaciated) will allow for a more rapid decrease of temperature (Spitz and Fischer, 1993; Vass et al., 2008).

Livor Mortis. Livor mortis, also known as lividity, occurs when the cells break down and circulatory activity ceases, causing red or purple areas on the skin where blood pooled in response to gravity and the positioning of the body at the time of death (Baden and Hennessee, 1989). When pressure is applied from objects such as clothing, blood does not pool in those locations and no discoloration is seen. Lividity begins approximately 30 minutes to one hour postmortem and becomes fixed at approximately eight to twelve hours after death (Mayer, 2012). Prior to this, blanching lividity is noticeable where light pressure on the livid area will leave a lighter imprint for a short time. After that, lividity is considered fixed and no blanching occurs in response to movement or pressure (Perper, 1993). The amount of lividity visible depends on the amount of blood present in the body (Mayer, 2012). In cases where there is trauma that causes blood loss, there will be less blood available to settle in the body, thereby leading to the decreased visibility or intensity of lividity when compared to an individual without such blood loss.

Each of these three processes, rigor mortis, algor mortis, and livor mortis, can be used to estimate a postmortem interval that is generally accurate to within a few hours. However, as decomposition progresses, the efficacy of these processes decreases and a need arises to utilize other techniques to estimate the postmortem interval such as the amount of discoloration, insect activity, or skeletonization.

Advanced Decomposition

Putrefaction. The increase in bacterial activity that leads to the further destruction of the surrounding tissues is known as putrefaction. This process, which is characterized by gross external changes to the body, begins approximately 48-72 hours postmortem (Janaway, 1996). During putrefaction, a decrease in oxygen results in the proliferation of internal bacteria that leads to the continued breakdown of the soft tissues and organs, and produces a buildup of internal gasses (Coe, 1993; Macchiarelli and Feola, 1995; Gill-King, 1997; Marks et al., 2009). Due to the fact that most of the digestive bacteria are located in the intestines, much of the color change and external signs of putrefaction, such as bloating and marbling, are first visible in the abdominal area and then spread to other areas of the body as the bacteria travel throughout the tissues (Love and Marks, 2003). Marbling is one of the early signs of putrefaction and is characterized by a breakdown of hemoglobin in the blood, leading to widespread purple, brown, black, or green venous discoloration (Love and Marks, 2003). As the body decomposes, a distinct odor is produced due to the breakdown of organic compounds (Mayer, 2012).

As the bacteria continue to consume the tissues, they excrete gases and the buildup of these gasses causes the abdomen to distend and become swollen (Marks et al., 2009). Although it has been stated that limbs do not bloat (Megyesi, 2001; Megyesi et al., 2005), distention and bloating can be observed in the arms and legs as well (Suckling, 2011). As bloating continues, the limbs can become splayed or bent and the body can change position or roll/lean completely toward one side (Smith, 2015). Putrefaction continues with the bloating in the body subsiding, purging of the body fluids, intense soft tissue deterioration, and desiccation and disintegration of the skin (Rodriguez and Bass, 1983; Galloway et al., 1989; Marks et al., 2009; Mayer, 2012). Once most of the gases have been expelled and the tissue continues to break down, bloating ceases. Sometime prior to the complete cessation of bloating, initial skeletonization may be observed, especially in the face or other areas of sustained insect activity (Love and Marks, 2003). Putrefaction, like other stages of decomposition, is significantly affected by environmental conditions such that moisture, temperature, and other bacteria can all affect the duration of this process (Janaway, 1996).

Adipocere Formation. Another component seen during putrefaction is the formation of adipocere, which develops from the breakdown of lipids during the process of saponification (Fielder and Graw, 2003; Ubelaker and Zarenko, 2011). In moist warm environments with little to no oxygen, adipose or fat tissue is converted into a solid material comprised mostly of saturated and unsaturated fatty acids. These fatty acids can then bind with sodium that is present in interstitial fluids or with potassium found in cell water, and form two types of adipocere. The first type occurs from sodium binding with fatty acids, rapidly forming adipocere that has a hard and crumbly texture. The second type develops from the interaction of potassium and fatty acids and slowly forms into a soft, paste-like consistency (Forbes, Stuart, Dardour, and Dent, 2004; Wilson-Taylor, 2013). While adipocere can be found in a variety of individuals, it is more commonly found in individuals with a higher percentage of body fat such as the obese, infants, and women (Ubelaker and Zarenko, 2011; Wilson-Taylor, 2013). While the formation of adipocere may complicate PMI estimations due to its preservative properties, it can be useful in forensic investigations because it can also preserve perimortem injuries and toxicological evidence (Ubelaker and Zarenko, 2011)

Decay. This last stage is characterized by the increased and intense soft tissue decomposition that leads to the further breakdown and disintegration of the tissues by bacteria and insects, and the initial exposure of bone in localized areas (Rodriguez and Bass, 1983; Galloway et al., 1989; Sledzik, 1998). During decay, the soft tissues become increasingly exposed to the environment, which allows for oxygen to enter and increases bacterial activity, resulting in accelerated tissue decomposition (Rodriguez and Bass, 1983; Sledzik, 1998). If hair and scalp remain on the skull, they will slough off and form a hair mat, a mass of hair that collects beneath the head. Once the bodily fluids have purged from the body, a process started in putrefaction, the remains begin to dry out, leading to further bone exposure and eventual skeletonization.

Late Decomposition

Skeletonization. The final stage of the decomposition process involves extensive skeletonization of the entire body with little to no soft tissue present (Rodriguez and Bass, 1983; Galloway et al., 1989). This is the longest lasting stage, beginning when at least fifty percent of the skeleton is exposed and persisting until the eventual breakdown of the bones themselves (Megyesi et al., 2005). It is with skeletonization that the influence of taphonomic factors (e.g., weathering, plant and animal involvement) can lead to much variability in the appearance and condition of the remains. Determination of the postmortem interval can be difficult once a body has reached skeletonization because bone quality and appearance will depend more on the environmental circumstances than the decomposition process.

Mummification. Another outcome in late-stage decomposition is mummification, or the dehydration of soft tissues that results in the stiffening of the skin, leading to preservation of the body (Galloway et al., 1989; Catanese and Bloom, 2002; Janaway, Percival, and Wilson, 2009). At this stage, insect activity is very limited and the body is no longer considered to be actively decomposing (Catanese and Bloom, 2002; Marks et al., 2009). Mummification results from the body's exposure to dry, arid environments (Galloway et al., 1989; Janaway et al., 2009), when the body is clothed and/or wrapped in cotton blankets (Dautartas, 2009), or when larvae feed on the internal tissues, leaving the skin intact to dry out and mummify. It is possible for rehydration of mummified remains to occur if they are exposed to heavy rain or other moisture that then rehydrates the desiccated tissues and invites the recolonization of insects (Suckling, 2011; Suckling, Spradley, and Godde, 2015). Due to these factors, postmortem interval estimates made from mummified remains are very tenuous because the body enters a period of stasis where little discernable change occurs over time.

VARIABLES AFFECTING DECOMPOSITION

There is a plethora of factors that can influence the decomposition process at any stage by either accelerating the process and promoting insect activity, or retarding decomposition and hindering insect development. The most widely discussed factors include temperature (Galloway et al., 1989; Mann, Bass, and Meadows, 1990; Janaway, 1996), humidity (Smith, 1984; Galloway et al., 1989; Aturaliya and Lukasewyz, 1999), insect and animal access (Haglund, Reay, and Swindler, 1988; Campobasso, DiVella, and Introna, 2001; Srnka, 2003; Klippel and Synstelien, 2007; Synstelien, 2015), depositional context (Mant, 1987; Janaway, 1996; Janaway et al., 2009), and individual variation, including size (Mant, 1987), pathological conditions (Polson, 1996), and drug/alcohol use (Dayananda and Kiran, 2013). While these variables will be discussed separately, their effects do not occur in a vacuum since multiple factors affect decomposition simultaneously. Because of this, it is difficult to isolate a single cause for a particular decomposition sequence and this highlights the variable nature of this process.

Temperature

Temperature is considered the most important factor influencing decomposition and in fact, most other influential factors are temperature dependent (Galloway et al., 1989; Mann et al., 1990; Marks et al., 2009). Typically, decomposition is accelerated, bacterial activity is promoted, and insect activity is pronounced in warmer months, while the opposite is true in colder temperatures: decomposition slows and there is a decrease in insect activity (Rodriguez and Bass, 1983; Smith, 1984; Mann et al., 1990; Janaway, 1996; Adlam and Simmons, 2007); although extremely high temperatures may inhibit bacterial proliferation and encourage mummification instead (Galloway et al., 1989; Micozzi, 1997). Likewise, cold environments, where temperatures remain at or below freezing, can significantly prolong decomposition and act as an agent of preservation by discouraging both insect and scavenger activity (Micozzi, 1991; Janaway, 1996). Tangentially related to temperature is the relationship between the amount of sun and shade at the depositional location. Bodies placed in full sun are more likely to experience desiccation and mummification of the soft tissues due to warmer temperatures and UV radiation, than bodies placed in the shade which is typically a cooler, moister environment and therefore, will slow down the decomposition process and lead to less desiccation and mummification (Galloway et al., 1989; Janaway, 1996).

Humidity

Like ambient temperature, humidity is considered a critical factor in decomposition. Typically, humid climates promote decomposition, while arid ones tend to slow down the process, resulting in mummification, especially when combined with higher temperatures (Smith, 1984; Galloway et al., 1989; Aturaliya and Lukasewyz, 1999; Janaway et al., 2009). The combination of heat and aridity can be particularly likely to lead to the mummification of remains. As noted previously, covering or wrapping the body in material that absorbs body moisture promotes desiccation while materials that trap moisture on the body and do not promote evaporation tend to accelerate decomposition and prevent mummification (Aturaliya and Lukasewyz, 1999; Dautartas, 2009).

Insect and Animal Access

In addition to climatological characteristics, other features of the environment play a large role in the decomposition process, such as insect and animal scavengers, by promoting decomposition and accelerating skeletonization.

Insects. Blowflies are typically the first insects observed on the body, arriving just minutes after deposition (Campobasso et al., 2001). They typically lay eggs in and around the warm, dark, moist areas of the body, including the mouth, nostrils, eyes, anal-genital area, skin folds, and open wounds. These eggs then hatch into larvae (maggots), within eight to fourteen hours, depending on ambient temperature (Campobasso et al., 2001). Once hatched, maggots burrow into the soft tissue, leading to skin slippage and the eventual liquefaction of the areas due to the digestive enzymes and the heat produced by the insects during these activities (Evans, 1963; Lord, 1990; Mann et al., 1990).

Insect activity, including oviposition timing and colony size, are greatly dependent on ambient temperature and sun exposure. Warmer, sunnier areas tend to promote faster insect colonization, but the direct sunlight restricts the colony's size (Srnka, 2003). Cooler, shadier spots will result in slower colonization rates but larger populations, due to the lack of insects' exposure to direct sunlight (Srnka, 2003)

Animal Scavengers. Non-insect scavengers can also affect the rate of decomposition. Scavenging is an important aspect of the decomposition process especially when it occurs during the fresh, early, or advanced stages, where it can have the greatest effect on postmortem interval estimation. Much research has been done describing the effects of scavenging on remains in a variety of scenarios, both on soft tissue and bone (Haynes, 1982; Haglund et al. 1988; Galloway et al., 1989; Haglund, Reay, and Swindler, 1989; Haglund, 1997a,b; Komar and Beattie, 1998;

Berryman, 2002; Klippel and Synstelien, 2007; Steadman and Worne, 2007; Reeves, 2009; Ricketts, 2013; Smith, 2015; Suckling et al., 2015).

Many different species of animal have been found to scavenge from human remains. Canid scavenging is one of the most well documented types of scavenging in terms of their effects on bone, disarticulation patterns, dispersal patterns of skeletal elements, and the relocation of elements (Willey and Snyder, 1989; Haglund, 1997a). Large carnivores such as wolves or dogs are primarily responsible for consuming the soft tissues of the abdomen, face, and neck and then disarticulating limbs from the torso (Haglund et al., 1989; Willey and Snyder, 1989; Haglund, 1997a,b). Other scavengers such as rodents (e.g., squirrels and rats) typically do not consume the soft tissue, but instead gnaw on the shafts and epiphyses of longbones in order to obtain minerals such as calcium or fats such as those found in the yellow marrow at the epiphyses of longbones (Haglund et al., 1989; Klippel and Synstelien, 2007).

Another scavenger that is responsible for extensive scavenging and scattering of remains is the raccoon. Northern raccoons (*Procyon lotor*) are quite ubiquitous across North America and are found in both urban and rural locales due to their highly adaptable nature (Roussere et al., 2003; Beasley, DeVault, Retamosa, and Rhodes Jr., 2007; Beasley, DeVault, and Rhodes Jr., 2007; Souza, Ramsey, Patton, and New, 2009; Synstelien, 2015). Raccoons are opportunistic scavengers and very resourceful in their scavenging habits (Beasley DeVault, Retamosa, et al., 2007; Beasley, DeVault, and Rhodes Jr., 2007; Souza et al., 2009). Their advantage lies not only in their intelligence, but also in their use of their forepaws, which have five toes that can be used to grasp and manipulate objects similar to a human hand (Halfpenny and Bruchac, 2002; Synstelien, 2015).

The effect of scavengers on the decomposition process can be extensive. Depending on the amount of scavenging, remains can go from fresh to partial skeletonization in a matter of hours. This can inflate PMI estimations because the body appears to be in a later stage of decomposition and hence older, simply due to scavenging. This can then lead to misidentifications or no identification being made due to an inaccurate timeline.

Depositional Context

Buried vs. Surface Placement. Whether remains are deposited on the surface or buried also greatly affects how they decompose. The burial of remains inhibits both insects and scavengers, thereby delaying decomposition, and temperatures above ground are generally higher than below the surface, resulting in a slower process underground (Mant, 1987; Janaway, 1996; Fielder and Graw, 2003). Burial also makes detection more difficult, resulting in longer internment times and possibly more extensive decomposition prior to discovery and recovery.

Individual Variation

Body Size. Decomposition can also be affected by characteristics of the deceased individual (Stuart, 2003). Variables such as obesity and fat composition, for example, can slow the rate of postmortem cooling as subcutaneous and abdominal fat provide insulation to the body's tissues (Zhou and Byard, 2011). A higher fat percentage may also lead to adipocere formation which in turn, promotes preservation of soft tissues and hence a slower decomposition process (Clark, Wornell, and Pless, 1997). Mant (1987) discussed the differences in individuals with higher versus lower body fat percentages, where lower body fat individuals tended to skeletonize more rapidly than individuals with higher body fat proportions. Larger individuals required a much longer time for skeletonization to occur because excessive body fat can reduce the dissipation of heat, which is an essential component for decomposition (Gonzales, Vance, Helpern, Umberger, 1954; Mant, 1987). In addition, body fat encourages bacterial growth, a necessary component of putrefaction (Mant, 1987). Thus, although bodies with a higher fat content will begin to decompose quickly, the overall process leading to skeletonization will typically require more time than individuals with lower fat content. Furthermore, individuals with more muscle tissue may decompose slower than those individuals with a greater fat percentage because they are not as 'insulated' from the fat tissue. In addition, skeletal muscle and connective tissues are some of the most tenacious tissues and are some of the last to decompose (Gill-King, 1997).

Pathological Conditions and Trauma. Antemortem infections can contribute to accelerated putrefaction, as microbial agents are already active within the body prior to the start of decomposition (Polson, 1996). Furthermore, decomposition is accelerated from infectious diseases due to the combination of increased antemortem bacteria, preexistence of bacteria in the organs and blood, and the probability of an elevated body temperature (Polson, 1996). Putrefaction may be increased in tissues in case of renal or congestive heart failure because bacteria are more effectively spread through fluids like blood (Zhou and Bayard, 2011). Trauma to the body, resulting in open wounds, can accelerate decomposition, mostly through insects' attraction to the areas of trauma. Insects typically oviposit in natural orifices; however, open wounds provide easier opportunities for feeding and egg deposition (Mann et al., 1990).

Drugs and Alcohol. The presence of drugs and/or alcohol in the body at the time of death may affect insect development and the decomposition process (Dayananda and Kiran, 2013). Antibiotics taken during life, especially just prior to death, can alter the number of bacterial colonies present in the gut. As bacterial proliferation is critical for the onset of putrefaction, a decrease in the type and quantity of bacteria can delay this stage of decomposition because it would take longer for the bacteria to proliferate than if the individual died with an average amount of gut bacteria (Dayananda and Kiran, 2013). Alternatively, drugs such as cocaine can increase the rate of decomposition due to its effects on the insect life cycle. According to the field of entomotoxicology, when fly larvae ingest tissues that contain stimulants such as cocaine, they experience an increased rate of growth resulting in a larger body size and faster movement (Dayananda and Kiran, 2013). This increase in size and activity level results in an increased rate of decomposition as the larvae are quickly consuming the soft tissues.

ESTIMATING THE POSTMORTEM INTERVAL

The accurate and reliable estimation of the postmortem interval is the ultimate goal in understanding variable decomposition rates. Due to the numerous factors affecting decomposition, narrow PMI estimations are oftentimes difficult to achieve. The early stages of decomposition provide more narrow and precise timeframes because the changes that occur within the first 24-48 hours after death are more predictable than later stage changes. For example, rigor mortis occurs approximately two to four hours after death, peaks at eight to twelve hours, and begins to recede within 24 hours, whereas bloating seen in late stage decomposition, can occur anywhere from days to months after death, depending on the precise environmental conditions (Janaway, 1996; DiMiao and DiMaio, 2001; Marks et al., 2009). Therefore, it is necessary to develop more refined methods of estimating time. One such method that has become more prevalent as a standardized measure of PMI is accumulated degree-days.

Accumulated Degree Days (ADD)

Accumulated degree-days, or ADD, are average daily temperatures used as a proxy for chronological time. ADD has its origins in entomology where a need existed for estimating time in the insect life cycle, which is greatly dependent upon temperature (Anderson, 2000). ADD are calculated by first finding the average daily temperature from the high and low temperature of each day and then summing the average temperatures over consecutive days (Miller, 2002; Megyesi et al., 2005). If the temperature is at or below zero Celsius on any day, rather than negative values, a value of zero is used (Miller, 2002). For example, if the average daily temperatures over three days were 15°C, 5°C and -2°C, the ADD would be calculated as 15+5+0, resulting in an ADD value of 20. Vass, Bass, Wolt, Foss, and Ammons (1992) showed that volatile fatty acid production, the byproducts of soft tissue breakdown, ceased around 1285 +/- 110 ADD, corresponding to the decay and skeletonization stages, thereby discovering the length of time required for the completion of active decomposition. This value is applied to cases where the number of days a body has been decomposing is unknown. Vass et al. (1992) recommended calculating the average daily temperatures over one week and dividing that into 1285 with the resulting value being an estimate of the maximum number of days since death. Because this method is based on the average time it takes a body to cease volatile fatty acid production, it better approximates the actual time since death for bodies that are in advanced or late stages of decomposition (Vass et al., 1992; Miller, 2002). However, in cases where the body has not reached advanced decomposition or where volatile fatty acid production is ongoing, this method may not accurately estimate time since death and other methods that rely on decomposition characteristics, like the Total Body Score method described below (Megyesi et al., 2005) must be utilized.

ASSESSING DECOMPOSITION

Total Body Score (TBS)

While ADD is a useful method for estimating time, the assessment of a body's decomposition characteristics is just as important for PMI estimation. One method of assessing decomposition is the Total Body Score (TBS) system developed by Megyesi (2001). This assessment assigns a numerical value to the observed stage of decomposition for three body segments: 1) head and neck, 2) trunk, and 3) limbs (Megyesi, 2001; Megyesi et al., 2005). Each of the three body segments is assigned a score based on the amount and type of decomposition observed that corresponds to a particular category: Fresh, Early Decomposition, Advanced Decomposition, and Skeletonization (Table 2; Megyesi et al., 2005). The three body segment scores are then summed to obtain the Total Body Score (TBS) that represents the total amount of decomposition present over the entire body. For example, as shown in Table 2 if all body segments are observed to be fresh with no discoloration or other signs of decomposition, each segment would be scored as a one (1) and the TBS would equal three (3). One advantage to this assessment is that by dividing the body into segments, greater detail can be achieved than if the body was examined holistically. This breakdown can help pinpoint areas where decomposition may be progressing faster or slower than other body segments and postmortem interval estimates can be modified to take that into account. However, the Total Body Score system was developed on 68 cases of complete photographs of human remains in a variety of settings (e.g. sun, shade, indoors, covered, etc.). Because of this limited cross-sectional sample, TBS assessment does not include many aspects of decomposition including bloating of the limbs, mummification without the presence of bone exposure, skin discoloration in the red, yellow, or orange spectrum, or accelerated decomposition due to scavenging. Even though photographs were the basis for the TBS system, its use has been expanded to include longitudinal field applications with either humans (Bytheway et al., in press; Dautartas, 2009) or pigs (Myburgh et al., 2013). Daily observation of the donor and assessment of TBS may exacerbate these subjective evaluations and

Table 2: Categories of decomposition for the head/neck, trunk, and limbs (Megyesi et al. 2005).

Categories and Corresponding Descriptions for Scoring the Head & Neck	**Assigned Score**
A. Fresh	
1. Fresh, no discoloration.	1 pt
B. Early Decomposition	
1. Pink-White appearance with skin slippage and some hair loss.	2 pts
2. Gray to green discoloration: some flesh still relatively fresh.	3 pts
3. Discoloration and /or brownish shades particularly at edges, drying of nose, ears, and lips.	4 pts
4. Purging of decompositional fluids out of eyes, ears, nose, mouth, some bloating of neck and face may be present.	5 pts
5. Brown to black discoloration of flesh.	6 pts
C. Advanced decomposition	
1. Caving in of the flesh and tissues of eyes and throat.	7 pts
2. Moist decomposition with bone exposure less than one half that of the area being scored.	8 pts
3. Mummification with bone exposure less than one half that of the area being scored.	9 pts
D. Skeletonization	
1. Bone exposure of more than half of the area being scored with greasy substances and decomposed tissue.	10 pts
2. Bone exposure of more than half of the area being scored with desiccated or mummified tissue.	11 pts
3. Bones largely dry, but retaining some grease.	12 pts
4. Dry bone.	13 pts
Categories and Corresponding Descriptions for Scoring the Trunk	**Assigned Score**
A. **Fresh**	
1. Fresh, no discoloration.	1 pt
B. **Early Decomposition**	
1. Pink-white appearance with skin slippage and marbling present.	2 pts
2. Gray to green discoloration: some flesh still relatively fresh.	3 pts
3. Bloating with green discoloration and purging of decompositional fluids.	4 pts
4. Postbloating following release of the abdominal gases, with discoloration changing from green to black.	5 pts
C. **Advanced Decomposition**	
1. Decomposition of tissue producing sagging of flesh; caving in of the abdominal cavity	6 pts
2. Moist decomposition with bone exposure less than one half that of the area being scored.	7 pts
3. Mummification with bone exposure less than one half that of the area being scored.	8 pts
D. Skeletonization	
1. Bones with decomposed tissue, sometimes with body fluids and grease still present.	9 pts
2. Bones with desiccated or mummified tissue covering less than one half of the area being scored.	10 pts
3. Bones largely dry, but retaining some grease.	11 pts
4. Dry bone.	12 pts
Categories and Corresponding Descriptions for Scoring the Limbs	**Assigned Score**
A. Fresh	
1. Fresh, no discoloration.	1 pt
B. Early Decomposition	
1. Pink-white appearance with skin slippage and marbling present.	2 pts
2. Gray to green discoloration: some flesh still relatively fresh.	3 pts
3. Bloating with green discoloration and purging of decompositional fluids.	4 pts
4. Postbloating following release of the abdominal gases, with discoloration changing from green to black.	5 pts
C. Advanced Decomposition	
1. Decomposition of tissue producing sagging of flesh; caving in of the abdominal cavity	6 pts
2. Moist decomposition with bone exposure less than one half that of the area being scored.	7 pts
3. Mummification with bone exposure less than one half that of the area being scored.	8 pts
D. Skeletonization	
1. Bones with decomposed tissue, sometimes with body fluids and grease still present.	9 pts
2. Bones with desiccated or mummified tissue covering less than one half of the area being scored.	10 pts
3. Bones largely dry, but retaining some grease.	11 pts
4. Dry bone.	12 pts

lead examiners to rely on scores or comparisons to decomposition stages from the last observation point. This then requires subjectivity on the part of the observer to determine what characteristics are present and which numerical value to assign to each segment. This can be problematic when, for example, the arms and legs are observed to be in two different categories or two individuals located in close proximity display much different decomposition sequences. This can produce ambiguity and uncertainty for the observer. Such uncertainty is beneficial for this current study, however, because the introduction of ambiguity may cause participants to rely on heuristics, biases, or other cognitive shortcuts to resolve that ambiguity Examining TBS in both longitudinal and cross-sectional applications will highlight the observers' use of these cognitions in different conditions.

Variations on TBS

While Megyesi et al.'s (2005) scoring system is an important first step in standardizing decomposition assessment, it has its limitations and several recent studies have addressed these concerns. Nawrocka, Fratczak, and Matuszewski (2016) examined the degree of inter-rater reliability of the TBS system in assessing photographs of pig carcasses. Their goal was to demonstrate that the TBS method has reliability when used by raters with different levels of experiences and that individuals outside of the anthropological community could benefit from its use. Overall, Nawrocka et al. (2016) found that the TBS system produced homogenous results regardless of the background of the rater. Additionally, the authors found that error rates increased with more advanced decomposition stages such that scores were less accurate for remains in advanced decomposition than for earlier stages. Typically, scores were overestimated in early stages and underestimated in later stages. Aside from the suggestion of elaborating on or rewording some descriptors to make them less ambiguous, one area of improvement for this study is the authors' application of the TBS to non-human remains. The TBS system was developed using human subjects (Megyesi et al., 2005) and while often used in forensic research as proxies for humans, pigs are not adequate for this system as they do not have the same body proportions or structure as humans and insects or scavengers may interact with pig carcasses differently from their interaction with human remains (Dautartas et al., 2018).

Dabbs, Connor, and Bytheway (2016) conducted a similar study that utilized photographs of 16 human remains. They found, similar to Nawrocka et al. (2016), that TBS has high levels of reliability and they therefore eliminated the scoring method as a source of error in PMI estimations. The errors they did find were related to user errors in calculating the TBS or in correctly labeling the category chosen, recording a value of 'C7' for example, instead of seven points. The areas that had lower reliability include the trunk and limbs. These issues mainly involved users being unable to differentiate between pre-bloat gravity-related skin sagging in the trunk and post-bloat sagging in photographs, and the wide range of coloration in the trunk that is not included in the TBS system. The issues with the limbs revolved around the limbs decomposing at different rates, making scoring difficult.

CONCLUSION

The definitions and stages described and discussed in this chapter were all utilized in the scoring and analyses of human decomposition made during the course of this study. Throughout the history of forensic anthropology, understanding the sequence of decomposition and the many variables that can influence the progression from one stage to the next, has been of great importance, especially in its application to postmortem interval estimation (Bass, 1997; Clarke et al., 1997; Marks et al., 2009; Megyesi et al., 2005). As discussed in this chapter, many variables can influence the decomposition sequence. However, as described in Chapter 2, very few studies have examined the variables that can influence the examiner. The prevalence of cognitive bias has only recently been examined within the forensic sciences and no studies up to this point have yet examined the role of bias within human decomposition analyses, especially focusing on the subjectivity and ambiguity present in the Total Body Score system. This study attempts to fill that gap in the literature by assessing how observers' perceptions, motivations, and mood states influenced their ability to accurately judge decompositional stages and how those biases might affect the estimation of the postmortem interval. Chapter 4 details the materials, procedures, and analyses used to assess those biases. The results of these analyses will be discussed in Chapter 5.

CHAPTER FOUR:

MATERIALS AND METHODS

MATERIALS

As discussed in Chapters 2 and 3, cognitions, such as perceptions, motivations, and affect, play a role in the decision-making process. This study examines cognitive biases within forensic anthropological methods, specifically the Total Body Score system, utilized in human decomposition research (Megyesi et al., 2005; Dror and Fraser-Mackenzie, 2009). Analysis of the decompositional and cognitive data began with establishing a standard for accuracy between the TBS values of experts utilizing Cronbach's alpha (α) as well as comparing that standard to observers' scores via hierarchical linear modeling. These tests were used to address the first hypothesis of this dissertation that there would be a difference between experts and observers' TBS values. To analyze this difference further, hierarchical random intercept multiple regression models were conducted on the data obtained from the observers' answers to the IMI and PANAS questionnaires in order to assess their ability to explain TBS values. This methodology is also used to address hypothesis 2, that the differences detected between groups would be due to the cognitive variables of motivation and mood. Lastly, hierarchical linear models were again used to examine photograph assessments as described in hypothesis 3, where it is expected that individuals will produce more accurate TBS values from scoring a donor in-person than from photographs.

Human Subjects

The data collected in this study consisted of body segment values, total body scores (TBS), IMI and PANAS subscale values, and responses from short answer questions provided to

54 University of Tennessee-affiliated students (50 observers, 4 experts) during both field and photograph assessments. This study evolved out of a National Institutes of Justice-funded study in which the donors used for this study were provided by grant DOJ-NIJ-2014-DN-BX-K009. The observers included 48 undergraduate volunteer assistants and two graduate students who were selected for participation in this project based upon their initial interest and involvement with the Forensic Anthropology Center (FAC); experts consisted of four graduate students with extensive experience in human decomposition. Observers were excluded from this study if they did not meet the requirements for working at the FAC (e.g., current Biosafety Training issued through the University of Tennessee's Office of Biosafety) or if they discontinued their involvement in the study. The University of Tennessee's Office of Research approved the IRB protocol and materials for this project on 27 June 2016 (UTK IRB-16-0309XP, renewed 12 June 2017) and immediately prior to their initial scoring session, participants read and signed an informed consent form. To ensure the anonymity of responses, a three-digit identification number was chosen by each observer at the beginning of the study, which they then recorded on their data packet at each scoring session for the duration of the project.

Of the 50 total observers, 41 completed the field assessment and 36 completed the photograph assessment. These samples include 24 observers who completed both phases and 12 who only completed the photograph assessment; experts completed both phases. All observers who scored at the ARF were trained via a lecture on the basics of the decomposition process and on the use of the TBS system. This was done to ensure that observers had some knowledge of decomposition prior to scoring because the TBS system references specific processes such as marbling, skin slippage, and mummification (Megyesi et al., 2005). The same training was attempted for the photograph assessment, however, only three of the 12 observers participated due to an email complication where some potential observers did not receive the training schedule and were therefore unable to complete training. This small number is possibly attributed to the training sessions occurring close to the end of the semester where individuals may not have had the time to devote to it. Therefore, the remaining untrained observers were allowed to complete the photo-scoring task and all reported no confusion or difficulty in using the scoring system or with any of the terms used to describe the decomposition characteristics.

Experts

In order to assess whether observers' TBS and body segment scores were accurate, a discussion of accuracy is required. Since the TBS system is reliant upon observers' judgments regarding the presence of certain characteristics that fit Megyesi's (2001) categories, assessment of accuracy is difficult. It is therefore not enough to state that the scores of a single person are "accurate" as they are prone to the same potential biases as the observers. A better option is to have a group of individuals who are experienced and knowledgeable in the decomposition process score the same donors as the observers. If the scores of these 'experts' are compared and found to be statistically similar, then that will serve as the standard for accuracy against which the observers' scores will be compared.

Four graduate students served as experts for this study. These graduate students, identified as Experts 1, 2, 3, and 4, were chosen on the basis of their experience with decomposition. Each of them were responsible for daily photographs at the ARF and some have also served as graduate research assistants for decomposition-related projects, thus, they all have been exposed to the different manifestations of decomposition seen in outdoor environments. Between them, they have over six years of experience working with human decomposition. Expert 1 scored twice a day in the morning and afternoon corresponding to the observers' scoring sessions, while Experts 2, 3, and 4 alternated scoring the donors for one-week intervals in both Trial 1 (Fall) and Trial 2 (Winter/Spring). Therefore, there were two experts per day for accuracy determination. In the rare occurrence that an expert missed a day of scoring (two occasions), their scores from the previous and following days were averaged and substituted for the missing data. These experts did not undergo training due to their previous familiarity with decomposition and the scoring system.

Outcome Measures: IMI and PANAS

As described in Chapter 2, the Intrinsic Motivation Inventory (IMI) and the Positive and Negative Affect Schedule (PANAS) were chosen for inclusion in this study in order to measure observers' motivations and moods as these are the main cognitive factors hypothesized to have an affect on scoring accuracy. The experts did not complete the IMI or PANAS due to time constraints: they were completing their daily responsibilities on top of scoring for this study, so the added scales were seen as too much of a daily time investment for them.

Intrinsic Motivation Inventory (IMI). The IMI is a self-report measure that assesses an individual's subjective experience while performing a task, which in this study is the analyses of decomposition characteristics via the TBS method. While the full IMI is composed of seven subscales representing the constructs of interest/enjoyment, perceived competence, effort, feelings of pressure/tension, value/usefulness, perceived choice during the task, and relatedness, only the first four were included in this study because they are hypothesized to be the main explanatory factors driving decision-making in the observers (Table 3; Appendix A). This subset of the scales was chosen because they best represent the motivations expected to influence observers' scoring decisions. For example, the relatedness subscale measures interpersonal interactions and friendship formation and as this task did not include an interpersonal component,

Table 3: The Intrinsic Motivation Inventory items used in the study

For each of the following statements, please indicate how true it is for you, using the following scale:

1	2	3	4	5	6	7
not at all true			somewhat true			very true

Score provided by subject	**Statements**
	I enjoyed doing this activity very much
	I think I am pretty good at this activity.
	I put a lot of effort into this.
	I did not feel nervous at all while doing this.
	After working at this activity for a while, I felt pretty competent.
	This activity was fun to do.
	I felt pressured while doing these.
	I am satisfied with my performance at this task.
	I didn’t try very hard to do well at this activity.
	I was very relaxed in doing these.
	I was pretty skilled at this activity.
	While I was doing this activity, I was thinking about how much I enjoyed it.
	It was important to me to do well at this task.
	I was anxious while working on this task.
	I felt this task got more difficult over time.
	I think I did pretty well at this activity, compared to other students.
	I didn’t put much energy into this.
	This activity did not hold my attention at all.
	I felt very tense while doing this activity.
	This task was very tedious.
	I tried very hard on this activity.
	I would describe this activity as very interesting.
	This was an activity that I couldn’t do very well.
	I thought this activity was quite enjoyable.
	This task was easy to do.

this scale was left out of the IMI for this study. The advantage to using a subset of the IMI is not only a reduction in observer fatigue, but it allows for a more efficient focus on only those items that relate to the goals of this study. A great advantage of the IMI is the ability to include only those subscales relevant to the particular study parameters. Ryan (1982) found a negligible effect of item order and no impact on scale validity with the inclusion or exclusion of specific subscales. In fact, it is rare that all subscale items have been used in a single experiment (McAuley et al., 1987). For these reasons, the IMI was considered ideal and appropriate for this study.

Additional Items. In addition to the four subscales of the IMI, three items were created, tested, and added into the IMI to assess observers' perceived difficulty with the scoring task (Table 3; Appendix A). The scale consisted of the items "This task was easy to do," "I felt this task got more difficult over time," and "This task was very tedious." The items had good reliability ($\alpha = 0.89$) and low intercorrelations with the four IMI subscales used in this study, ranging from -0.12 with the pressure/tension subscale to 0.47 with perceived competence.

Positive and Negative Affect Schedule (PANAS). The PANAS is a self-report survey composed of 20 emotion-related traits (10 positive, 10 negative) to assess how much observers identify with each emotion at a specific time period (Table 4). For this study, the instructions indicated for observers to "indicate to what extent you feel this way right now, at the present moment" because it is hypothesized that mood relates to the observers' level of accuracy and mood at the time of completing the scoring task is of interest. Similar to the IMI, both the positive and negative affect subscales of the PANAS have been demonstrated to be robust with good internal consistency, reliability, and validity (Tellegan, 1985; Watson and Clark, 1984; Watson et al., 1988). However, unlike the IMI, the individual items of the PANAS cannot be altered or excluded without reassessing the scale's psychometric properties. Therefore, this study

Table 4: The Positive and Negative Affect Schedule (PANAS)

This scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you feel this way right now, that is, at the present moment. Use the following scale to record your answers:

1	2	3	4	5
very slightly or not at all	a little	moderately	quite a bit	extremely

Score provided by subject	**Positive Affect**	**Score provided by subject**	**Negative Affect**
	Interested		Irritable
	Determined		Upset
	Excited		Ashamed
	Inspired		Disinterested
	Strong		Nervous
	Attentive		Guilty
	Alert		Scared
	Active		Jittery
	Enthusiastic		Hostile
	Proud		Afraid

utilized all 20 original items of the PANAS without alteration (Table 4; Appendix A).

Additional Outcome Measures

After signing the informed consent form, all participants were provided with a brief packet that included the scoring materials and survey items (Appendix A). The scoring materials included a scoring sheet with spaces designated for each body segment score and the Total Body Score (TBS) complete with a copy of Megyesi et al.'s (2005) scoring rubric, which provides a description of the decomposition characteristics corresponding to each numerical value (Table 2). As described in Chapter 3, the TBS system assigns a numerical value to the observed characteristics of decomposition for three body segments: head and neck, the trunk, and the limbs (Megyesi, 2001; Megyesi et al., 2005); these three scores are summed to obtain the total amount of decomposition present on the entire body (the Total Body Score). The questionnaire items contained after the scoring materials included both the IMI and PANAS scales (Tables 3 and 4) and demographic questions asking about sex, age, year in school, role within the FAC, and any prior experience and/or knowledge of decomposition (Appendix A). Additionally, participants were also asked to describe which decomposition characteristics they viewed to be the most important when making their scoring choices, and any additional comments they had (e.g., the ease/difficulty of using the TBS system) in both the field and photo phases (Appendix A).

METHODS

Phase One: Field Assessment

Of the 50 total observers, 41 provided Total Body Scores (TBS) for ten donated individuals (six in Trial 1, four in Trial 2; Table 5) placed at the ARF, encompassing the mid-Fall to mid-Spring seasons; the remaining 12 participated only in the Phase II photo assessment. The different seasons provided observers with examples of decomposition in various environmental conditions. All donors selected for this study were free from trauma or autopsy, weighed between 58.97 kg and 113.4 kg, placed supine, and were contained in wooden cages measuring 3'x 3' x 7' to reduce scavenger activity (as dictated by the NIJ-sponsored study from which this project originated) while still allowing access for insect activity (Table 5; Figure 1). Daily photos were taken of all donors and a selection of photos was used in the second phase (photograph scoring) of this project.

Each of the 41 observers provided scores for the three body segments for each donor in either the morning or afternoon, no more than twice a week, and not on consecutive days. The limitation on the number and timing of scoring sessions per observer per week was to minimize potential biases from repetition and familiarity with the scoring system and questionnaires. It was

Table 5: Demographic information of donated individuals used in this study.

Donor ID	Season	Age	Sex	Height (cm)	Weight (lb)	Placement Location
1	Fall	55	M	177.8	160	Mix Sun/Shade
2	Fall	61	F	167.6	150	Shade
3	Fall	72	M	-------	--------	Mix Sun/Shade
4	Fall	63	M	167.6	220	Sun
5	Fall	71	F	167.6	103	Shade
6	Fall	77	M	-------	200	Sun
7	Winter	77	M	170.2	212	Sun
8	Winter	78	F	167.6	165	Shade
9	Winter	74	F	157.5	200	Sun
10	Winter	82	M	172.7	180	Mix Sun/Shade

Figure 1: Example of cages used to reduce scavenger activity.

important that observers did not remember or utilize their scores from previous sessions when making their decisions, however, this could not be guaranteed. Additionally, observers could choose to score in the morning or afternoons. The choice of two scoring times was made to not only provide more opportunities for observers to participate, but also to determine if time of day had any influence on scoring. Daily fluctuations in temperature during the fall and spring months could alter characteristics such as the visibility of bone exposure: increased insect activity during the warmer afternoons sometimes obscured bone exposure that was visible in the cooler mornings. Temperature and weather conditions could affect observers' moods and motivations and thus lower accuracy. For example, extreme temperatures or rain may decrease observers' attentiveness to the scoring task so they could rapidly finish and move indoors.

After scoring each of the body segments and obtaining a TBS for each set of remains, observers completed a series of questionnaires aimed at understanding their decision-making processes and included the IMI and PANAS (Appendix A). Scoring concluded for a donor once bone exposure was observed in each body segment (corresponding to at least stage C2 for head/neck and trunk and C1 for limbs; Megyesi et al., 2005; Table 2; Figure 2). Once that amount of bone exposure occurred, scoring continued for one week in order for all observers to score that stage, before scoring was discontinued for that individual. Neither Trial 1 nor Trial 2 had all donors reach bone exposure in all three segments. Two of the six donors in Trial 1 exited the study due to extensive bone exposure from scavenging; none of the four donors in Trial 2 reached that amount of bone exposure. Therefore, Trial 1 lasted for seven weeks, from 15 October 2016 – 2 December 2016, while Trial 2 continued for 10 weeks, from 20 February 2017 – 30 April 2017. The end dates correspond to the end of the academic terms.

Figure 2: Donor #5 displaying bone exposure in all three body segments. Arrows indicate areas of bone exposure.

Phase Two: Photograph Assessment

The Total Body Score system was originally developed using photographs from forensic cases (Megyesi, 2001). Since its development, it has been utilized in both photograph and field assessments with the latter incorporating both cross-sectional and longitudinal research (Bytheway et al., in press; Dabbs et al., 2017). Because photographs and fieldwork are essential elements of forensic anthropological casework, both settings were studied here.

During Trial 2, observers were asked to use the TBS system to score a series of 20 photographs consisting of 16 photos of donors from both trials and four photos of donors unfamiliar to the observers. Of the 20 photos, 14 consisted of full body images where all three body segments were visible, while six depicted isolated segments (two each of head/neck, trunk, legs; Figures 3-6). Body segment scores are to be made using only the information that can be obtained from the individual segments. That is, the head and neck are not to influence the scores

Figure 3: Full body image used in Phase II photo scoring task.

Figure 4: Head/neck only image used in Phase II photo scoring task.


Figure 5: Trunk image used in Phase II photo scoring task.


Figure 6: Limbs image used in Phase II photo scoring task.

for the trunk or limbs. However, when the entire body is observable, either in the field or in full body photographs observers may be unknowingly using their knowledge of one segment to influence their judgment of the others. For example, if the limbs show evidence of drying and darkening, but the trunk is not bloated and still has areas of fresh tissue, then observers may feel compelled to score the limbs at an earlier stage of decomposition due to their knowledge of the trunk's condition. This would help them make sense of the donor's decomposition pattern and timeline if all of the segments corresponded to a similar timeframe. In the field, this is unrealistic to control, as all segments are visible. However, photographs can easily be edited to focus on a single segment (Figures 4-6). In Phase II then, observers were instructed to only score those segments they believed contained enough information for them to score. For example, in Figure 5, only the trunk is fully visible, but part of the arms, thighs, and neck are also present in the photo. If observers believe they have enough information to score all of those segments, then it elucidates their possible decision-making process and how they approach the TBS method.

Identical to the field assessment, observers scoring from photographs completed a series of questionnaires about their decision-making process, motivations, and mood (Appendix A). A sample of 12 observers who did not score in the field assessment also completed the photograph-scoring task in order to control for any repetition biases that might linger in the other sample. As mentioned previously, only three of these new observers completed the same decomposition training and instruction with the scoring system as those who participated in the field assessment. However, no difficulties with the task were reported.

STATISTICAL ANALYSES

All TBS and questionnaire data were entered into an Excel spreadsheet and imported into SPSS software, Version 23 for Mac (IBM, Armonk, NY, 2016) for statistical analyses. While every effort was made to limit the presence of missing data, factors such as poor weather, especially during Trial 2, and participant reliability could not be avoided. There were no days in Trial 1 that had 100% missing data and only six days where data were missing from either the morning or afternoon session. However, this does not imply that all observers came to all of their scheduled scoring sessions. For example, several observers missed a session, but data were collected from other scheduled observers that day. Observers who missed a session tended to reschedule to maintain the same number of sessions a week. Trial 2, however, had four days of completely missing data due to the Spring Break holiday where undergraduate observers were unavailable as well as several days of poor weather, which made data collection impossible. These missing data were noted in the spreadsheet and excluded from the analyses.

Accuracy Assessment

As discussed previously in Chapter 3, establishing accuracy within the TBS system is problematic because the system requires a human observer to make a judgment regarding how the observed decomposition characteristics fit into the scoring system; there is no current method for determining correctness of a score because too much subjectivity is involved. Therefore, the TBS and body segment score standards are established by the experts (see *Experts* section in this chapter) in order to obtain the 'accurate' TBS and segment scores. Consistency within the experts' scores was assessed with Cronbach's alpha (α), a coefficient of reliability that measures internal consistency, which is whether a set of items that propose to measure the same general construct produce similar scores. In this case, it is assessing how closely related the experts scores are to each other. This is done in order to set a standard with which to compare the observers' data. An alpha coefficient value of 0.7 or higher is considered the threshold for determining acceptable internal consistency (Kline, 2000; DeVellis, 2012) and that threshold was used in these analyses.

Following the establishment of expert consistency, the accuracy of observers' TBS scores was accomplished by comparing their scores with those of the experts to determine if a difference existed between these groups. Due to the dependence of the data from the same observers scoring the same donors over time, a hierarchical linear model (HLM) with random intercept was utilized to account for the dependence and variation in scores due to donors, group (experts and observers), time, and body segments. HLM utilizes maximum likelihood estimation (MLE) to estimate the fixed and random factors to examine variance when the predictor variables represent different levels (Corbeil and Searle, 1976; Boedeker, 2017). For example, the 'group' variable contains two levels, those of experts and observers. Significant differences (p<0.05) would provide support for Hypothesis 1 that differences exist between observers and experts, and these differences may be related to internal or external factors, such as those measured by the IMI and PANAS.

To examine differences in accuracy further, multiple regression was used to understand the relationship between the IMI and PANAS subscales and TBS values via a hierarchical random intercept model. Furthermore, in this regression model, predictor significance (measured as $\alpha < 0.05$) as well as the meeting of normality assumptions are used to determine model fit.

Field versus Photographs

As this study was also designed to determine the effect, if any, the scoring medium – field versus photograph – had on accuracy, the differences between field and photograph trials were assessed. As in the field assessment, expert data were subjected to Cronbach's α to assess internal consistency within the 20 photographs. As in Phase I, hierarchical linear models were employed to compare experts and observers. These models included the same levels as in Phase I, group, body segment, and time, with the addition of a photograph variable representing the 20 individual photographs. Additionally, a separate model compared those who participated in both the field and photograph assessments with observers who only scored photographs to determine if field experience played a role in observers' scoring decisions. Similar to the field assessments, hierarchical random intercept multiple regression was utilized to assess the relationship between observers' scores and cognitive factors such as their self-reported motivation and mood as measured by the IMI and PANAS. It was expected that the field assessments will show greater accuracy than the photographs and that the latter will increase the subjectivity and resulting frustration due to the uncertainty and ambiguity each image presents. The ambiguity stems from the fact that photographs from a fixed distance limits available information to observers compared to field observers who can analyze the body in greater detail. The following chapter, Chapter 5, presents the results of this study.

CHAPTER FIVE:

RESULTS

The results presented in this chapter address whether cognitive biases affect observers' assessments of decomposition characteristics, both in the field and photographs, when compared to a group of experts. Phase I data for both Trial 1 and Trial 2 are presented to address Hypotheses 1 and 2, specifically whether or not observers' TBS values differ from those of experts and the role cognitive processes play in decision-making. Following this, Phase II results are presented that directly address Hypothesis 3, that experts and observers will score photographs differently due to the added constraints imposed by a two-dimensional image; the role of cognitions is also addressed within Phase II.

PHASE I: FIELD ASSESSMENT: TRIAL 1 (FALL)

Phase I compared the TBS data of 29 undergraduates and four graduate student experts collected from six donated individuals from 15 October 2016 – 2 December 2016 at the ARF.

Reliability

Expert 1. As discussed in Chapter Four, Expert 1 scored twice a day at each scoring event, while Experts 2-4 scored once a day on alternating weeks at the ARF. Because it was not consistently recorded at what time Experts 2, 3, and 4 conducted their scoring, a direct comparison with Expert 1's morning or afternoon scores was not possible. Therefore, Expert 1's body segment and TBS values would need to be averaged so as to compare them with the other experts as well as assess them for internal consistency. If those scores significantly differ between morning and afternoon sessions, then a single average would not be appropriate.

SPSS v. 23.0 (IBM, Armonk, NY, 2016) was used to run Cronbach's alpha (α) analyses to assess the consistency between Expert 1's morning and afternoon scores. Cronbach's α assesses the agreement or consistency between different measurements. In this case, do the morning and afternoon scores agree? The results for Expert 1 across all donors are presented in Table 6a. These results demonstrate a high internal consistency ($\alpha \geq 0.7$) between morning and afternoon sessions. If the data are examined by donor, the same pattern emerges where all values show high internal consistency (Table 6b). Internal consistency is concluded when the coefficient of alpha reaches 0.70 or higher (Kline, 2000; DeVellis, 2012). As shown in Table 6b, the range of alpha coefficients is 0.865 – 0.998, well above the 0.70 threshold. Therefore, it can be concluded that the scores from the morning and afternoon sessions are consistent and a daily average can be calculated for comparison with the daily scores of Experts 2-4.

Table 6a: Expert #1's reliability for morning and afternoon scores across all donors in Trial 1.

Head/Neck	Trunk	Limbs	TBS	N
0.993	0.981	0.978	0.986	210

Table 6b: Cronbach's α values for Expert #1 within all donors of Trial 1 between morning and afternoon sessions.

Donor	Head/Neck	Trunk	Limbs	Total Body Score	N
1	0.996	0.988	0.975	0.997	49
2	0.993	0.988	0.975	0.995	49
3	0.994	0.993	0.997	0.998	24
4	0.988	0.996	0.965	0.994	36
5	0.995	0.944	0.976	0.985	16
6	0.949	0.925	0.925	0.865	36

All Experts. As discussed in Chapter Four, it is important to demonstrate consistency of scores between the experts. If the experts are not sufficiently consistent, then they will not be able to function as the standard with which to compare the observers' data. Therefore, as with the data for Expert 1, Cronbach's α was utilized to assess the level of consistency between all four experts. Table 7a shows the α coefficients of the experts across all segments and donors. When further broken down by body segment, the high internal consistency result holds with α values ranging from 0.835 to 0.995 (Table 7b). Internal consistency was measured within each donor as well to check for any inconsistencies in scoring due to the individual differences in the donors. As shown in Table 7c, internal consistency between the four experts remained high within the six donors and this demonstrates that if the individual characteristics of the donors exerted any influence over the scores of the experts, all experts' scores were equally affected. Because of this internal consistency, it can be concluded that the experts do agree with regards to their body segment and TBS scores and these scores can be averaged for comparison with the data provided by the undergraduate observers.

Table 7a: Comparison of all experts across all donors/segments in Trial 1.

	Expert 2	Expert 3	Expert 4
Expert 1	0.991 (N=228)	0.982 (N=268)	0.983 (N=312)

Table 7b: Reliability (and sample size) of Expert 1 with Experts 2-4 within body segments in Trial 1.

Expert	Head/Neck	Trunk	Limbs	TBS	N
2	0.935	0.995	0.924	0.982	57
3	0.928	0.895	0.835	0.951	67
4	0.873	0.947	0.847	0.956	78

Table 7c: Reliability of Expert 1 with Experts 2-4 within donor for Trial 1.

Donor	Expert 2	N	Expert 3	N	Expert 4	N
1	0.986	56	0.989	56	0.987	76
2	0.992	56	0.994	56	0.993	76
3	0.994	28	0.984	28	0.966	36
4	0.991	36	0.983	52	0.969	52
5	0.971	8	0.944	24	0.990	28
6	0.991	36	0.973	52	0.976	52

Evaluating Observer Accuracy

Hypothesis 1 states that significant differences are expected in the body segment scores and TBS values between experts and observers. To address this hypothesis, a hierarchical linear model (HLM) was estimated to compare body segment scores in observers and experts. This random intercept model consisted of four levels representing donor, body segments, group (experts and observers), and time (Appendix B). Body segment (i.e., head/neck, trunk, and limbs) was included in the model because they could play an influential role in scoring results since decomposition may progress faster in one segment than others. For example, initial skeletonization usually occurs first in the head/neck region due to the presence of insect activity in that area. However, as the focus of this dissertation is on the observer, it is not necessarily important to document *how* each donor decomposed. The results of the HLM analysis revealed significant differences in the scores between the observers and experts as well as over time (Table 8). These results suggest that observers tended to provide lower TBS scores than experts across all segments and donors. Table 9 displays the means and mean differences between the expert and observer scores for the segments for each donor. Overall, the mean scores for observers were significantly different from those of the experts. For example, the mean score for the observers' TBS for donor five ($\bar{x}$ = 18.640) was significantly different than the experts' TBS for the same donor ($\bar{x}$ = 19.578; Table 9).

Table 8: Results of fixed effects of the HLM for Trial 1.

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Group	1	4578	4.18	0.0409
Day	1	5	15.95	0.0104
Day*Group	1	4578	3.24	0.0717

Table 9: Mean differences between the segment scores for experts and observers by donor.

Donor	Segment	Expert Mean	Observer Mean	Difference between groups
1	Head/Neck Trunk Limbs TBS N	8.464 5.091 4.431 17.987 49	8.439 4.970 4.700 18.048 214	0.025 0.121 -0.269 -0.061
2	Head/Neck Trunk Limbs TBS N	7.627 5.142 5.005 17.795 49	7.446 5.089 4.782 17.365 214	0.181 0.053 0.223 0.43
3	Head/Neck Trunk Limbs TBS N	6.281 5.958 5.234 15.596 24	5.944 5.550 4.852 16.348 119	0.337 0.408 0.382 -0.752
4	Head/Neck Trunk Limbs TBS N	6.881 4.437 4.659 15.979 36	6.692 4.567 4.511 15.740 157	0.189 -0.13 0.148 0.239
5	Head/Neck Trunk Limbs TBS N	6.531 6.015 7.031 19.578 16	6.056 5.721 6.878 18.640 83	0.475 0.294 0.153 0.938
6	Head/Neck Trunk Limbs TBS N	5.368 4.042 3.868 13.360 36	5.686 3.987 3.792 13.479 157	-0.318 0.055 0.076 -0.119

The model also indicated that scores were significantly different over time (Table 8). This makes sense as it is expected that body segment scores will increase over time as decomposition progresses. The data from Trial 1 suggests that experts tend to, on average, have higher scores than the observers. There are a few exceptions to this as seen with the limbs and TBS scores of Donor 1, the TBS of Donor 3, the trunk of Donor 4, and the head and TBS of Donor 6 (Table 9). These results might reflect a difficulty in scoring individuals who were scavenged, mummified, or whose decompositional sequence did not conform to the parameters of the TBS system.

Cognitive Influences

While experts did not complete the cognitive surveys, it is possible to assess whether a relationship existed between the observers' body segment and TBS values and their scores on the Intrinsic Motivation Inventory (IMI) and the Positive and Negative Affect Schedule (PANAS), utilizing hierarchical random intercept multiple regression. First, subscale scores for both the five IMI subscales and the two PANAS subscales were calculated for each observer. The IMI subscales quantify an individual's interest/enjoyment, perceived competence, pressure/tension felt during the task, effort put forth on the task, and task difficulty. The PANAS subscales of positive and negative affect measure positive and negative mood states, respectively.

Categorical variables of time (day), donor, and segment were included in the model as either random effects or covariates. This allowed for the control of the variance associated with these variables in order to get a clearer picture of the relationship of the cognitive variables and scores. For example, it is expected that TBS values will increase with time and that they will differ across donors due to individual differences amongst donors. Additionally, due to the strong effect of time on TBS values, each observers' individual scores were used rather than an average of all scores. This has the added benefit of accounting for those individual changes in motivation and mood that may be eliminated with the use of averages. The results of the regression (Table 10) indicated that two predictors accounted for 82% of the variance ($R^2 = 0.82$). It was found that perceived difficulty (IMIDifficulty) of the task ($F(1, 3729) = 14.808$, $p < 0.001$; $B = 0.163$) significantly predicted scores as did the amount of perceived pressure/tension (IMIPressTens) observer's felt while scoring ($F(1, 3729) = 12.781$, $p < 0.001$; $B = 0.201$; Table 10). B coefficients represent the unstandardized coefficient of the model and indicates that a one unit increase in the independent variable, in this case IMIDifficulty and IMIPressTens, leads to an increase of 0.201 and 0.163 in the dependent variable, TBS scores. This model displayed normality and good fit as shown in the histogram of the unstandardized residuals in Figure 7 (Westfall and Henning, 2013; Tabachnick and Fidell, 2013).

PHASE I: FIELD ASSESSMENT: TRIAL 2 (WINTER)

Trial 2 of the Phase I Field Assessment study consisted of the TBS data of 31 undergraduate observers as well as the four graduate student experts from Trial 1. This trial collected data from four donated individuals at the ARF and was ten weeks in duration, from 20 February 2017 – 30 April 2017, approximately three weeks longer than Trial 1.

Reliability

Expert 1. As in Trial 1, I scored twice a day during the morning and afternoon scoring sessions, while Experts 2-4 scored once a day. Similarly to Trial 1, the time at which Experts 2, 3, and 4 conducted their scoring was not consistently recorded and so a direct comparison with my morning or afternoon scores was not possible. Therefore, my body segment and TBS values needed to be assessed for internal consistency via Cronbach's α before comparing them to the scores of the other experts.

Table 10: Parameter estimates of multiple regression for Trial 1. Significant parameters are highlighted

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	9.328	.611	15.262	.000	8.130	10.526
IMIInterestEnjoy	-.053	.072	-.734	.463	-.195	.089
IMIEffortImport	.123	.097	1.261	.207	-.068	.314
IMIPressTens	.201	.056	3.575	.000	.091	.311
IMIComp	.075	.042	1.791	.073	-.007	.158
IMIDifficult	.163	.042	3.848	.000	.080	.246
PA	-.029	.075	-.385	.700	-.177	.119
NA	-.237	.221	-1.074	.283	-.670	.196
Day	.159	.004	41.029	.000	.151	.166
[Segment=1]	-7.653	.283	-27.014	.000	-8.208	-7.097
[Segment=2]	-9.411	.283	-33.220	.000	-9.966	-8.855
[Segment=3]	-9.522	.283	-33.613	.000	-10.078	-8.967
[Segment=4]	0[a]	.	.	.	.	.
[Donor=1]	3.776	.266	14.217	.000	3.256	4.297
[Donor=2]	3.090	.265	11.657	.000	2.570	3.610
[Donor=3]	3.545	.306	11.573	.000	2.944	4.145
[Donor=4]	2.218	.284	7.803	.000	1.660	2.775
[Donor=5]	7.556	.342	22.093	.000	6.886	8.227
[Donor=6]	0[a]	.	.	.	.	.

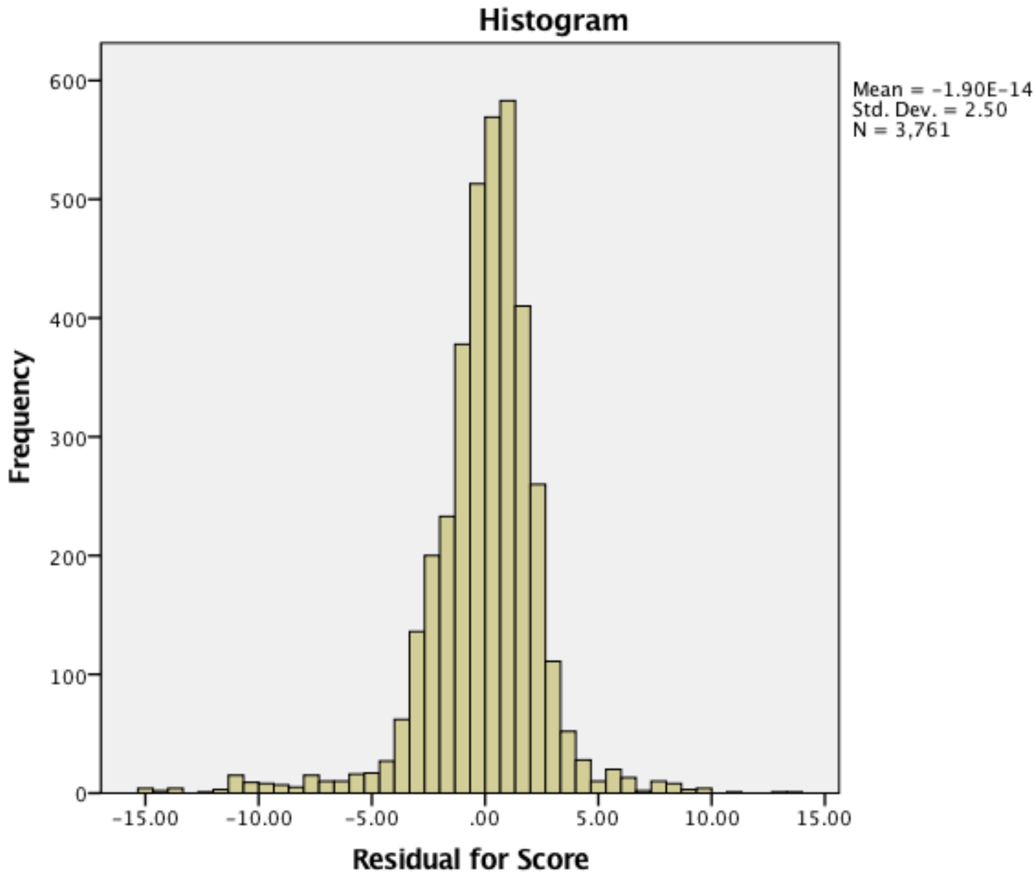

Figure 7: Histogram of the unstandardized residuals for Trial 1.

SPSS v. 23.0 (IBM, Armonk, NY, 2016) was used to run Cronbach's α analyses to assess the consistency between the morning and afternoon scores. These results across all donors are presented in Table 11a. Just like the results of Trial 1, high internal consistency was achieved between the morning and afternoons scores, both between and within donors (Tables 11a and 11b). As shown in Table 11b, the range of alpha coefficients is 0.977 – 0.998, well above the 0.70 threshold for good internal consistency (Kline, 2000; DeVellis, 2012). Therefore, it can be concluded that Expert 1's Trial 2 scores from the morning and afternoon sessions are consistent with one another and a daily average can be calculated for comparison with the daily scores of Experts 2-4.

Table 11a: Expert #1's reliability for morning and afternoon scores across all donors in Trial 2.

Head/Neck	Trunk	Limbs	TBS	N
0.994	0.992	0.986	0.997	212

Table 11b: Cronbach's α values for Expert #1 within all donors of Trial 2 between morning and afternoon sessions.

Donor	Head/Neck	Trunk	Limbs	Total Body Score	N
7	0.991	0.994	0.983	0.996	53
8	0.996	0.988	0.991	0.998	53
9	0.996	0.994	0.987	0.997	53
10	0.989	0.991	0.977	0.996	53

All Experts. Cronbach's α was again utilized to assess the level of consistency between all four experts. Table 12a shows the α coefficients of the experts across all segments and donors. When further examined by body segment as in Table 12b, the high internal consistency result holds for most experts. However, Expert 4 has an $\alpha = 0.696$ for the limb segment, which is the lowest α value obtained for both Trials 1 and 2 (Table 12b). While 0.696 falls just slightly below the threshold for good internal consistency, it was deemed close enough to 0.70 and none of

Table 12a: Comparison of all experts across all donors/segments in Trial 2.

	Expert 2	Expert 3	Expert 4
Expert 1	0.981 (N=400)	0.972 (N=352)	0.985 (N=320)

Table 12b: Reliability (and sample size) of Expert 1 with Experts 2-4 within body segments in Trial 2.

Expert	Head/Neck	Trunk	Limbs	TBS
2 (N=100)	0.933	0.943	0.943	0.972
3 (N=88)	0.917	0.938	0.877	0.938
4 (N=80)	0.928	0.939	0.696	0.931

Table 12c: Reliability of Expert 1 with Experts 2-4 within donor for Trial 2.

Donor	Expert 2 (N=100)	Expert 3 (N=88)	Expert 4 (N=80)
7	0.986	0.977	0.988
8	0.977	0.973	0.992
9	0.972	0.971	0.983
10	0.992	0.973	0.982

Expert 4's other results (Table 12a-12c) displayed low alpha coefficients, so their data remained in the analysis. High internal consistency was also seen within donor (Table 12c) and therefore, like Trial 1, any effect of donor from individuals differences did not cause difficulty in scoring – all experts adapted to donor differences in the same way. Because of this internal consistency, it can be concluded that the experts do agree with regards to their body segment and TBS scores and these scores can be averaged for comparison with the data provided by the undergraduate observers.

Evaluating Observer Accuracy

As with Trial 1, Trial 2 seeks to address Hypothesis 1, which states that significant differences are expected in the body segment scores between experts and observers. Hierarchical linear models were constructed to address this hypothesis (Appendix B). This random intercept model had the same four levels as in Trial 1: donor, body segment, group (experts vs observers), and time (day). Significant effects were found between experts and observers as well as for time (Table 13).

Table 13: Results of the HLM for Trial 2.

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Group	1	3817	16.46	<.0001
Day	1	11	153.96	<.0001
Day*Group	1	3817	5.27	0.0217

Mean scores for the observers and experts demonstrate the opposite pattern as that found in Trial 1 with observers scoring higher, on average, than the experts across all segments and donors; Trial 1 observers scored lower on average than the experts (Table 14). For example, the mean of the experts' TBS for donor nine ($\bar{x}$ = 13.227) was significantly different than the observers TBS for the same donor ($\bar{x}$ = 15.072) and this pattern repeated across all segments and donors. Table 14 displays the means, and mean differences between the experts and observers scores for the segments for each donor.

A significant interaction between group and day was also found for Trial 2 (Table 13) indicating that the difference in scores between the experts and observers may be changing over time. A subsequent mixed effects ANOVA model was utilized to investigate how the scores for the observers and experts changed over the 10-week timespan of Trial 2 (Appendix B). Weeks were used instead of the daily data in the previous analysis, in order to make the pattern in the data more interpretable. Table 15 shows significant effects for group, week, and the interaction (group*week), indicating that the difference between the scores of the experts and observers changed over Trial 2. In other words, while observers always scored higher than experts, the magnitude of that difference changed from week to week. Specifically, as seen in Table 16 and

Table 14: Mean differences between the segment scores for experts and observers by donor for Trial 2.

Donor	Segment	Expert Mean	Observer Mean	Difference between groups
7	Head/Neck Trunk Limbs TBS N	6.521 4.071 3.966 14.566 70	6.775 4.199 4.382 15.368 282	-0.254 -0.127 -0.415 -0.801
8	Head/Neck Trunk Limbs TBS N	6.971 4.790 4.409 16.145 70	7.051 5.129 4.699 16.894 282	-0.080 -0.339 -0.290 -0.749
9	Head/Neck Trunk Limbs TBS N	6.239 3.425 3.574 13.227 70	6.501 4.330 4.215 15.072 282	-0.262 -0.905 -0.641 -1.845
10	Head/Neck Trunk Limbs TBS N	5.542 4.060 3.900 13.486 70	5.668 4.170 4.205 14.090 282	-0.125 -0.109 -0.305 -0.604

Table 15: Test of group*week interaction in Trial 2.

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Group	1	15	26.10	<.0001
week	9	5568	618.80	<.0001
Group*week	9	5568	3.00	0.0014

Figure 8, the interaction indicates that the two groups significantly differed only during week one and not during any other week in Trial 2. As an interaction was not found for Trial 1, the pattern seen here may be due to the cold temperatures encountered during Trial 2, which slowed the progression of decomposition. This then would result in the donors displaying few changes from week to week, and most certainly little to no change day to day, until temperatures increased and decomposition progressed further. This may have led to the scores becoming similar between groups as all participants were seeing such little change over several weeks. This pattern and directions for future study will be discussed further in Chapter 6.

The results for Trial 2 demonstrate a pattern opposite to those of Trial 1. Here, observers tended to score donors higher on average, than the experts and that relationship changed over the course of Trial 2. As in Trial 1, the observers' own motivations, perceptions, and moods are next examined.

Cognitive Influences

A hierarchical random intercept multiple regression was employed in order to assess whether a relationship existed between the observers' scores and their responses on the IMI and the PANAS. The five IMI subscales and the two subscales of the PANAS were calculated for each observer and each subscale was run in the model as a predictor. Individual raw scores for each observer were used as the dependent variable to fully take into account how mood and motivation affect the scores.

Again, as with Trial 1, variances for donor, segment, and day were controlled by including them into the model (Table 17). The results of the regression indicate that three predictors accounted for 81% of the variance ($R^2 = 0.81$). It was found that, just like Trial 1, perceived difficulty (IMIDifficult) ($F(1,4394) = 22.410$, $p < 0.001$) and pressure/tension

Table 16: Comparisons of scores between experts and observers for each week of Trial 2.

Week	Estimate	Standard Error	DF	t Value	Pr > |t|
1	1.3403	0.2198	5568	6.1	<.0001
2	0.7917	0.2241	5568	3.53	0.0787
3	0.5863	0.2213	5568	2.65	0.1370
4	0.1964	0.2552	5568	0.77	0.4415
5	0.6179	0.2212	5568	2.79	0.9922
6	0.4373	0.2240	5568	1.95	0.5059
7	0.1111	0.2202	5568	0.50	0.6141
8	0.5624	0.2271	5568	2.48	0.2758
9	0.3361	0.2250	5568	1.49	0.1354
10	0.8165	0.2240	5568	3.65	0.0812

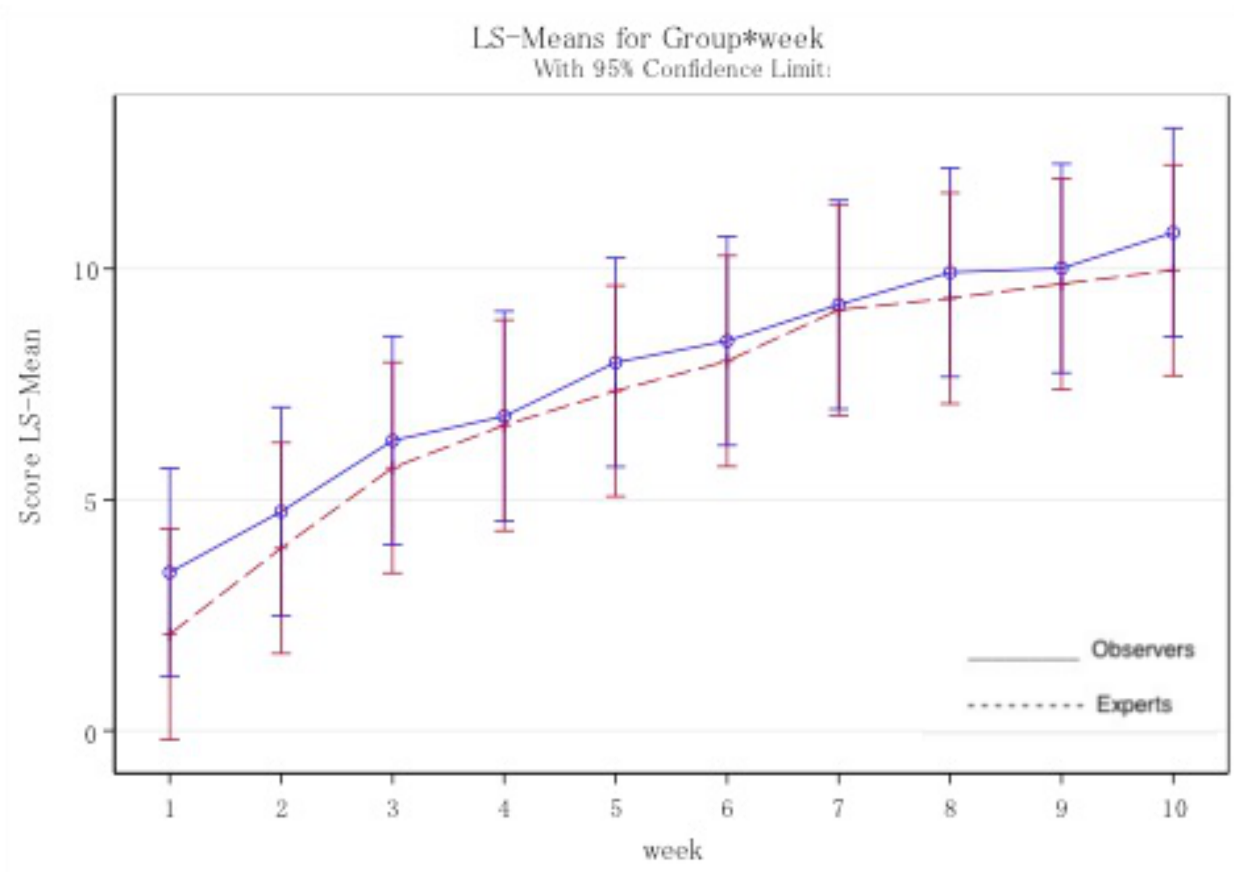

Figure 8: Graph of the weekly least square means and 95% confidence intervals of the experts and observers for Trial 2. Week one was the only week in which experts and observers significantly differed.

(IMIPressTens) subscale of the IMI (F(1,4394) = 5.915, p = 0.015) were significant predictors. However, Trial 2 also revealed that the influence of the competency IMI subscale (IMIComp), a measure of self-reported performance on the task, was also significant (F (1,4394) = 10.328, p = 0.001). For example, as feelings of competency increase, TBS scores increases by 0.110. And the same pattern occurs with difficulty. As observers view the task as difficult or tedious, the more likely their TBS and body segment scores increase. However, the opposite effect happens with pressure/tension: As feelings of anxiety, tension, and pressure increase for observers, TBS and body segment scores decrease by -0.122 (Table 17). Again, like Trial 1, mood appeared to have no significant influence on the scores in Trial 2. This model also shows good fit and normality as indicated in the histogram of the unstandardized residuals depicted in Figure 9.

Overall, during Phase I, perceptions of task difficulty, feelings of anxiety and pressure, and task competency appear to influence TBS scoring decisions. Surprisingly, neither positive nor negative mood states affected scores in either trial of Phase I. These results are echoed in the observers' responses to the question of "What characteristics did you use to help you choose your scores?" that was included in their post-scoring survey packet. While many observers responded with actual characteristics such as "discoloration," "skin colorization," "bloat," "absence of bloat," or "post bloat," many observers took the opportunity to comment on the task itself. For example, several commented that the color options and their timings listed in the TBS scale were limited or "out of order" in comparison to what they experienced in the field and this made scoring ambiguous and challenging. Consistent with the results of the regression analyses, some observers mentioned that their expectations or feelings influenced their scores. For example, one Trial 2 observer commented that she decided to score the trunk as a two for its" overall look, but still feeling uncomfortable assigning a 1" because the "trunk itself still

Table 17: Parameter estimates for Trial 2 multiple regression. Significant parameters are highlighted.

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	9.834	.479	20.510	.000	8.894	10.774
IMIInterestEnjoy	-.042	.048	-.876	.381	-.136	.052
IMIEffortImport	-.104	.082	-1.257	.209	-.265	.058
IMIPressTens	-.122	.050	-2.432	.015	-.221	-.024
IMIComp	.110	.034	3.214	.001	.043	.178
IMIDifficult	.146	.031	4.734	.000	.086	.206
PA	.082	.046	1.792	.073	-.008	.171
NA	-.075	.186	-.404	.686	-.441	.290
Day	.113	.001	75.745	.000	.110	.116
[Donor=7.0]	1.297	.172	7.547	.000	.960	1.634
[Donor=8.0]	2.801	.172	16.281	.000	2.464	3.139
[Donor=9.0]	.975	.172	5.671	.000	.638	1.312
[Donor=10.0]	0[a]	.	.	.	.	.
[Segment=1.0]	-8.394	.172	-48.872	.000	-8.730	-8.057
[Segment=2.0]	-9.884	.172	-57.554	.000	-10.221	-9.548
[Segment=3.0]	-9.838	.172	-57.281	.000	-10.174	-9.501
[Segment=4.0]	0[a]	.	.	.	.	.

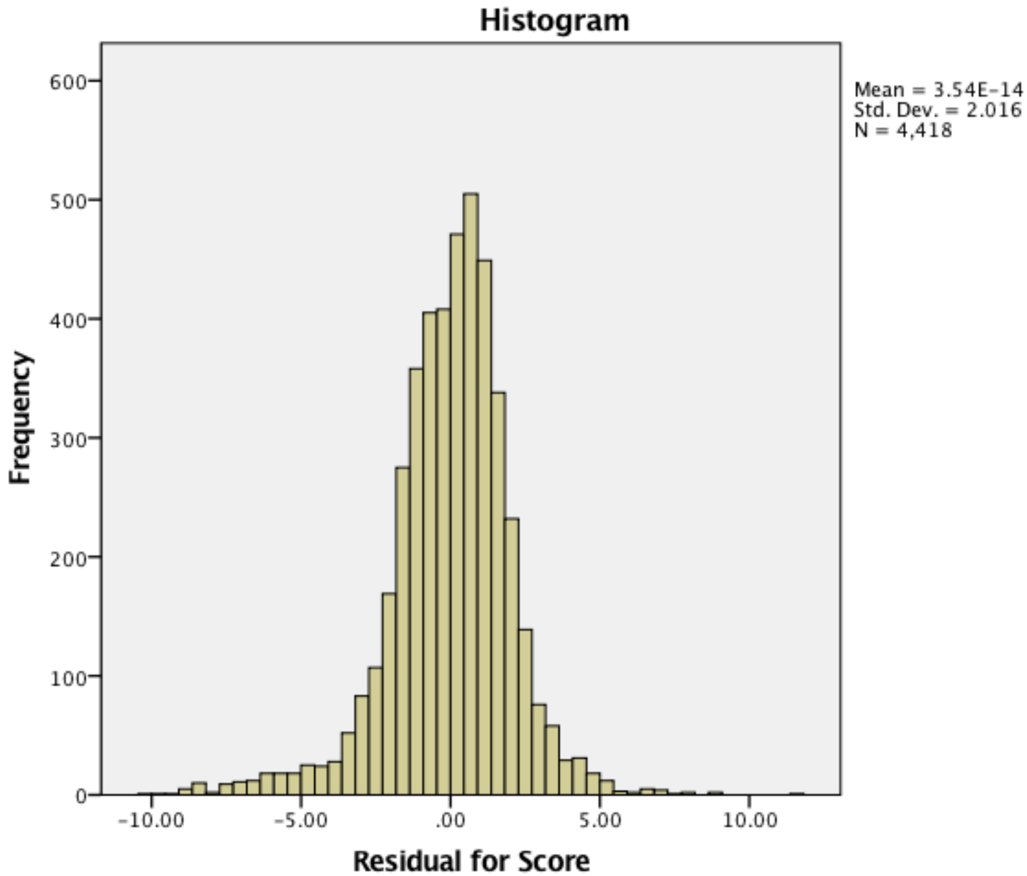

Figure 9: Histogram of the unstandardized residuals for Trial 2.

looks pretty fresh." Another Trial 2 observer mentioned that her decisions were "rushed" because there was a threat of rain, but she later commented that that same scoring session was the "first time it was extremely easy to follow the scoring sheet." That last statement regarding the ease or difficulty of the scoring task was only echoed by three other observers in the entirety of Phase I and it mostly occurred as the body transitioned from putrefaction to decay where the body dried out and the decomposition process slowed. This led to those observers commenting that the scoring task was "getting extremely hard" and it was "difficult to use the scoring system, especially as you get further along." These comments clarify the results obtained in Phase I and provide a contrast to the results obtained in Phase II.

PHASE II: PHOTOGRAPH ASSESSMENT

Phase II compared the TBS data of 35 observers and the four experts from Phase I for 20 photographs. Scoring commenced 11 April 2017 until 20 June 2017. As described in Chapter 4, these 20 photographs consisted of 14 full body images where all three body segments (i.e. head/neck, trunk, and limbs) were visible, and six photographs depicting isolated segments (two each of head/neck, trunk, and legs).

Reliability

Experts. As in the Phase I Field Assessment study, the four experts also completed the photo-scoring task. Cronbach's α was again utilized to assess the level of consistency between all four experts for the 20 photographs. Table 18a shows the α coefficients of the experts across all photographs while Table 18b shows the α coefficients for each of the 20 photographs. The scores of the experts across all photographs were internally consistent. Because of this, it can be concluded that, similar to Phase I, the experts in Phase II do agree with regards to their scores

Table 18a: Comparison of all experts across all photographs/segments in Phase II.

	Expert 2	Expert 3	Expert 4	N
Expert 1	0.991	0.991	0.988	80

Table 18b: Reliability of Experts 1-4 for Phase II. N=4 for all photos

Photograph	Experts
1	0.996
2	0.991
3	0.710
4	0.992
5	0.986
6	0.984
7	0.999
8	0.997
9	0.996
10	0.989
11	1.000
12	0.974
13	0.999
14	0.997
15	0.984
16	0.850
17	0.996
18	0.990
19	0.997
20	0.980

and these scores can be averaged for comparison with the data provided by the observers.

Evaluating Observer Accuracy

Phase II specifically addresses Hypothesis 3 which states that experts and observers will differ in their scores of the photographs. To address this hypothesis, observers and experts' TBS and body segment scores were analyzed via a mixed effects hierarchical linear model with repeated measures (Appendix B). This model was employed because each observer scored multiple photos and this model would account for the inherent relationship of the scores due to the same observers providing scores for all of the photographs. The model reveals no significant differences between experts and observers' scores on the photographs (Table 19).

Table 19: Results of mixed effects HLM for observers and experts in Phase II.

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Group	1	2047	0.31	0.5757
ARF(Group)	1	2051	0.35	0.5550

Because there were two different types of photographs present in this phase, full body and partial images, and this may influence observers' ability to score by presenting them with more or less information, another mixed effects model was utilized within these groups to see if the non-significance was due to type of photograph (Appendix B). The results indicate that observers and experts did not significantly differ in either the full body or partial body photographs (Table 20).

While no support was found for Hypothesis 3, this result may be due to the fact that the TBS system was developed utilizing photographs and therefore, the photos better fit the parameters of the TBS system and were not as ambiguous as assumed. These results will be

Table 20: Results of mixed effects HLM with ARF experience and photograph type included.

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Group	1	2061	0.30	0.5863
Photo_Type	1	2068	254.61	<.0001
ARF(Group)	1	2066	0.55	0.4594

further discussed in Chapter Six.

Accuracy and Experience

Because Phase II comprised both observers who had participated in Phase I and new observers with little to no experience with decomposition, this variable was included in the mixed effects models. As seen in Table 20, there is no significant difference in the scores for the Phase I observers who scored at the ARF and the Phase II only observers. These results, surprisingly, demonstrate that additional field experience had no effect on scores.

These results also show a significant effect photograph type (Table 20). This means that there appears to be a difference in scores due to type of photo, but that difference is not due to group membership, either as an expert/observer or prior field experience. More likely, this difference is due to the photographs depicting different donors and different body segments. Essentially, the value of the scores depends on the photo being scored. This, of course, makes sense because the photos display images with different body segments represented and it would be expected that scores would differ depending on the level of decomposition depicted in the image as well as the number of segments shown.

Cognitive Influences

As with Phase I, a goal of Phase II was to examine the influence motivation and mood exerted on observers' scores. Hierarchical random intercept multiple regressions were conducted on both the full and partial body photos. These photo types were analyzed separately because they represent separate challenges to the observers. While full body photographs depict all body segments, thereby allowing for all segments to be scored, the partial photographs depicted isolated segments and required observers to decide if enough information existed to score all or any, segments. There were six partial photographs included in Phase II, two each depicting only the head/neck, only the trunk, or only the legs. Overall, for five of the six partial photos, observers all chose the same segments to score. For example, all observers agreed that only the trunk segment contained enough information to be scored in Photo 6 (Figure 10); there was not enough information provided for the head/neck or limbs to be scored as the majority of those segments are not visible in the photo. However, Photo 20 (Figure 11) proved to be more ambiguous for observers. Of the 33 observers, 16 of them (48.48%) scored both the trunk and the limbs, while the remaining 17 observers (51.52%) only scored the trunk segment. It is interesting to note that the majority of those observers in the former category were those who had not been trained or been to the ARF before scoring these photos (9 of 12 observers). It may be that those with no prior ARF experience did not encounter the same ambiguity as did those observers who participated in Phase I. Therefore, while experience did not seem to affect the scores, that experience may have been important when deciding what to score in an ambiguous circumstance. However, this is only demonstrated on one photograph, so more research is needed before a conclusion can be reached.

As previously mentioned, different cognitive processes may be at work when analyzing full body versus partial body photos. The regression on the full body photographs revealed two predictors explaining 73% of the variance in scores ($R^2 = 0.73$; Table 21). Positive mood ($F(1, 1731) = 9.442$, $p = 0.002$) and perceived competency ($F(1, 1781) = 15.177$, $p < 0.001$) both

Figure 10: Photograph 6 of Phase II.

Figure 11: Photograph 20 from Phase II. While similar to photo 6 in the depiction of only part of the body, there were differences amongst observers in which segments were scored.

significantly predicted scores. This model, like those in Phase I, had good fit and normality, as shown in the histogram of unstandardized residuals field assessments (Figure 12).

The photos of partial bodies that focused on isolated segments were also subjected to a hierarchical random intercept regression with the same predictor variables as the fully body as well as controlling for the variance associated with photo and segment. While the full body photos showed positive affect and perceived competence as significant predictors of score, the partial photos were different. Only negative mood was a significant predictor of score ($F(1,203) = 3.902$, $p = 0.05$, $\beta = 0.744$). All other variables were not significant (Table 22). This model also had excellent normality and model fit as reflected histogram of the unstandardized residuals (Figure 13). These results are most likely due to the low sample size of the partial photos as only six photos depicted isolated segments while 14 were of the full body.

The fact that negative affect was only significant for partial photographs may reflect the overall frustration that observers experienced in trying to ascertain which segments to score and how to score them. For example, many of the observers commented that the partial photos were "frustrating," "SUPER hard," or " hella difficult." This may also hint subtly at their possible use and reliance on the context clues from other body segments to assign score.

As in Phase I, the observers also responded to the question of "What characteristics did you use to help you choose your scores?" and while observers did respond with actual characteristics, for example 60% mentioned coloration as their primary indicator of score, followed by bone exposure at 38%, an interesting pattern emerged between those observers who had participated in Phase I and those who had not. Phase I participants repeatedly mentioned that it was still very difficult to score the photos because they could not observe the different angles of the body like they could in the field and they felt that their scores were inaccurate. However,

Table 21: Parameter estimates of multiple regression for full body photographs from Phase II. Significant parameters are highlighted.

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Intercept	26.791	.704	38.077	.000	25.411	28.171
PA	.225	.073	3.073	.002	.081	.369
NA	.174	.202	.858	.391	-.223	.571
IMIIE	-.067	.064	-1.044	.296	-.192	.058
IMIEI	-.117	.138	-.849	.396	-.388	.154
IMIComp	.188	.048	3.896	.000	.093	.283
IMIPT	.001	.040	.027	.978	-.078	.080
IMIDifficulty	.033	.032	1.024	.306	-.030	.095
[Segment=1]	-16.318	.357	-45.693	.000	-17.019	-15.618
[Segment=2]	-17.621	.357	-49.342	.000	-18.322	-16.921
[Segment=3]	-20.455	.357	-57.275	.000	-21.155	-19.754
[Segment=4]	0[a]	.	.	.	.	.
[Photo=1.0]	-16.773	.357	-46.966	.000	-17.473	-16.072
[Photo=2.0]	-14.545	.357	-40.729	.000	-15.246	-13.845
[Photo=4.0]	-10.162	.360	-28.236	.000	-10.868	-9.457
[Photo=5.0]	-9.758	.357	-27.322	.000	-10.458	-9.057
[Photo=7.0]	-12.152	.357	-34.026	.000	-12.852	-11.451
[Photo=8.0]	-8.970	.357	-25.116	.000	-9.670	-8.269
[Photo=9.0]	-2.530	.357	-7.085	.000	-3.231	-1.830
[Photo=10.0]	-14.333	.357	-40.135	.000	-15.034	-13.633
[Photo=12.0]	-11.061	.357	-30.971	.000	-11.761	-10.360
[Photo=13.0]	-9.545	.357	-26.729	.000	-10.246	-8.845
[Photo=15.0]	-8.207	.360	-22.803	.000	-8.913	-7.501
[Photo=17.0]	-4.591	.357	-12.855	.000	-5.291	-3.890
[Photo=18.0]	-15.742	.357	-44.081	.000	-16.443	-15.042
[Photo=19.0]	0[a]	.	.	.	.	.

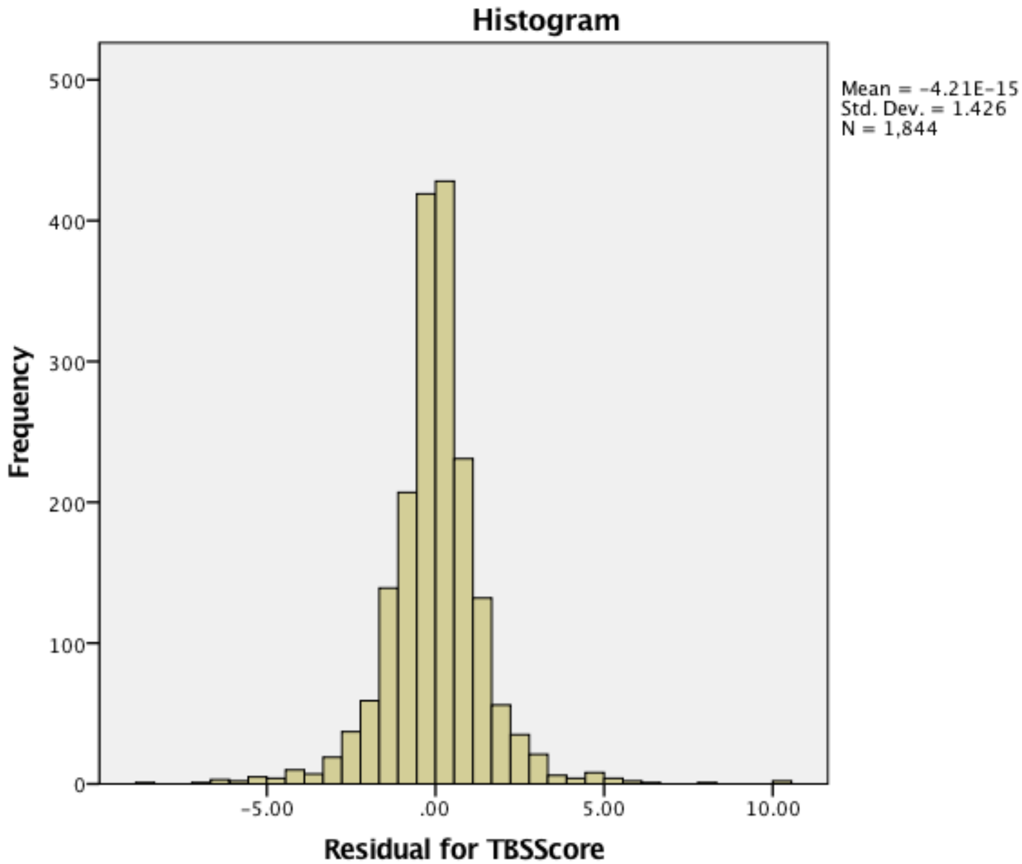

Figure 12: Histogram of the unstandardized residuals for Phase II full body photographs

Table 22: Parameter estimates for partial photos in Phase II.

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	.932	1.210	.770	.442	-1.453	3.317
PA	.255	.138	1.851	.066	-.017	.526
NA	.744	.377	1.975	.050	.001	1.487
IMIIE	-.165	.121	-1.365	.174	-.404	.073
IMIEI	.260	.255	1.020	.309	-.243	.763
IMIComp	.010	.090	.108	.914	-.167	.186
IMIPT	.094	.075	1.262	.208	-.053	.241
IMIDifficulty	-.066	.060	-1.101	.272	-.185	.052
[Segment=1]	-.114	.286	-.398	.691	-.678	.450
[Segment=2]	-.902	.286	-3.153	.002	-1.466	-.338
[Segment=3]	0[a]	.	.	.	.	.
[Photo=3.0]	-2.250	.286	-7.867	.000	-2.814	-1.686
[Photo=6.0]	-1.699	.967	-1.756	.081	-3.607	.208
[Photo=11.0]	1.455	.228	6.379	.000	1.005	1.904
[Photo=14.0]	2.364	.228	10.366	.000	1.914	2.813
[Photo=16.0]	0[a]	.	.	.	.	.
[Photo=20.0]	0[a]	.	.	.	.	.

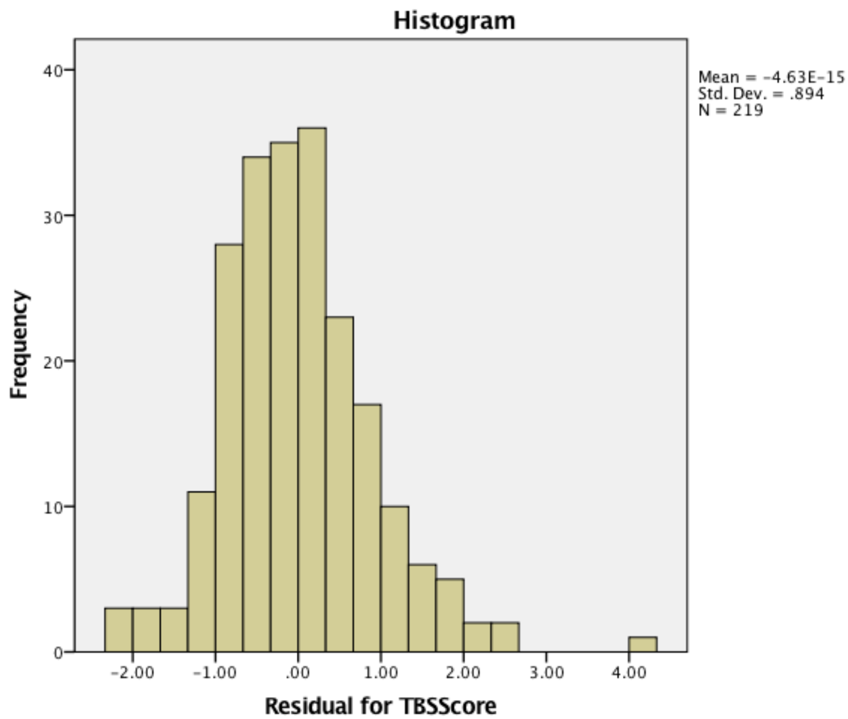

Figure 13: Histogram of the unstandardized residuals for partial body photos in Phase II

those observers who only participated in Phase II, and thus had no decomposition experience, stated that “the scoring was clear on what each value stood for,” “ it wasn’t difficult to find indicators…to help me score,” and that “there was nothing challenging about this.” Furthermore, 40% of these observers also stated that they thought the scoring system did not need to be improved and that it is “as good as it can be.” This differed dramatically from the Phase I observers who all had some suggestions for how the TBS system could be improved, particularly with its treatment of color.

A more complete discussion of these and all results in this chapter will be discussed in the next chapter, Chapter 6.

CHAPTER SIX:

DISCUSSION

The focus of this dissertation has been to evaluate whether cognitive biases alter the accuracy of decomposition analyses that use the Total Body Score system (Megyesi, 2001; Megyesi et al., 2005). Studying the progression of decomposition over time to estimate the postmortem interval is a fundamental component of forensic anthropology. In the current project, it was hypothesized that the observers will differ from experts in their scoring of decomposition characteristics because they have not acquired the knowledge and experience that experts have developed over multiple years (Dror, 2016; Dror and Charlton, 2016). It was further hypothesized that the cognitive measures of motivation and mood will be explanatory of observers' TBS values because these cognitive processes have been shown to effect decision-making abilities in other contexts (Ryan et al., 1990; Watson et al., 1998; Robbins and Judge, 2013). Finally, it was hypothesized that cross-sectional photograph-based TBS assessments will differ between experts and observers and it was also expected that mood and motivation variables would be significant predictors for the photographs scores. The results of each hypothesis will be discussed within the context that cognitive bias may affect an individual's ability to accurately assess decomposition, such that they cause total body scores to be over- or underestimated.

DIFFERENCES BETWEEN OBSERVERS AND EXPERTS

The results generated by the analyses for Hypothesis 1 revealed that both under- and overestimation of TBS occurred in Trial 1 and Trial 2, respectively. Those results may be due to a combination of biasing information and the different environmental factors present in each of the trials. For example, extensive, repetitive scavenging of two of the donors occurred throughout Trial 1. The impact of scavenging on remains is not included in Megyesi's (2001; Megyesi et al., 2005) system, so this required observers to adapt their scores to what they observed. Additionally, observers knew the approximate time since placement of each of the donors since they had been scoring them since placement at the ARF. This aspect of the study was not anticipated during the design, but may have had a large effect in retrospect. When trying to account for scavenging, observers may have underestimated the level of decomposition, especially as less soft tissue was present with which to use as the basis for their scores. It is reasonable to assume that since observers knew that Donor 5, for example, was only placed about 24-hours prior to her first scavenging event, that this information would influence their scores (Figures 14 and 15). In other words, observers may be succumbing to contextual or confirmation biases (Stelfox and Pease, 2005; Dror and Fraser-Mackenzie, 2009; Rossmo, 2009): they know that the donor appears to have characteristics typical of more advanced stages of decomposition (e.g. partial skeletonization) due to the external influence of scavengers, but they also know that the donor was only placed a few days prior. Therefore, they may have scored the donor in order to better reflect this timeframe, rather than based strictly on the characteristics and as a result, they tended to overcorrect for scavenging.

The results of Trial 2 show that observers tended to overestimate their scores when compared to experts. This overestimation may be due to both the prior knowledge of the donors' time since placement as well as seasonality. Trial 2 occurred mostly during the winter season, where below freezing temperatures were present through mid-March, approximately two-thirds of the duration of the trial. Combined with the fact that observers knew the approximate time


Figure 14: Donor #5 on the day of placement prior to scavenging.


Figure 15: Donor #5 after multiple scavenging events on both arms, legs, and the pelvis. Days since placement = 7.

since placement, as in Trial 1, this may have led to confirmation bias. Cold temperatures slow down the progression of decomposition (Micozzi, 1991; Janaway, 1996), leading to the persistence of the characteristics of fresh/early decomposition for several weeks. Therefore, observers may assume that the donors were "older" than their appearance would indicate due to the cold temperatures and so they scored them as such, leading to overestimation when compared to the experts' scores. These results are underscored by the fact that all observers were instructed, sometimes repeatedly, to score each donor as if they had never seen them before, therefore the scores would be independent of prior sessions. However, it is not easy to forget that information and so the placement timeframe may have always been on their mind and, consciously or not, influencing their results.

The implications from the results of Trials 1 and 2 highlight the influence that expectations have on the decision-making process; this influence is greater in ambiguous or challenging situations such as the ones here (Dror and Fraser-Mackenzie, 2009; Coutts and Gruman, 2012; Robbins and Judge, 2013). As mentioned above, in Trial 2, observers knew the donors' placement date and they were aware from their training that cold temperatures tended to slow down the decomposition process (Mann et al., 1990; Micozzi, 1991; Janaway, 1996; Adlam and Simmons, 2007). They then may have sought out any information or characteristics that confirmed that timeframe, while downgrading the importance of the fresher appearance of the body. This is typical of confirmation bias (Stelfox and Pease, 2005; Dror and Fraser-Mackenzie, 2009; Rossmo, 2009) and aligns with the results found by Nakhaeizadeh, Hanson et al. (2014). In that study, erroneous contextual information regarding the age at death of skeletal remains significantly impacted the estimates provided by participants. When given information that a set of remains was of a young adult (25-30 years old), participants provided lower age estimates than those participants who were informed that the remains were from an older adult (50-55 years old; Nakhaeizadeh, Hanson et al., 2014). In a similar way, observers in the current study utilized the contextual information of placement time to provide higher scores that corresponded to the donors' placement timeframe, rather than their overall appearance. This was also observed in Trial 1 where placement time was influential in the scoring of scavenged donors. In this case, as the bodies became skeletonized after only a few days due to scavenging, observers provided lower body segment scores than the experts, thereby emphasizing the importance of the short timeframe over the appearance of the remains.

While observers tended to score higher than experts throughout Trial 2, the overestimation effect was found to be statistically significant only for the first week of the study. This first week is the participants' first experience of scoring these particular donors and it represents a new scenario for their application of the TBS method; the groups' scores may differ initially due to this. Furthermore, the experts are experienced with the decomposition process and they, like the observers, knew placement information, so they may already have preconceived expectations of what the donors will look like on any given week. They may therefore, only consider certain TBS categories when they are scoring. For example, in the first week, the experts knew the donors would most likely all fall into the "fresh" category (Table 2) and so they do not consider anything beyond that. Observers, however, are much less experienced and may be considering all categories at first, hence their divergence from the experts. Additionally, Trial 2 took place from 19 February to 30 April 2017 and the first four weeks of the study had temperatures ranging from -5.8°C to 24.3°C (21.5°F to 75.7°F) with an average temperature during this time of 10.4°C (50.7°F). It was not until week five when temperatures increased for the duration of Trial 2, ranging from 3.6°C to 32.3°C (38.5°F to

90.2°F) and an average temperature of 18.5°C (65.3°F). This temperature difference is important because the progression of decomposition slows and insect activity is severely diminished in temperatures at or below 10°C (50°F; Galloway et al., 1989; Mann et al., 1990; Janaway, 1996; Marks et al., 2007). Therefore, decomposition progressed slowly for the majority of Trial 2 resulting in the donors displaying few visible signs of decomposition from week to week and when changes did occur (e.g. appearance of characteristics signaling advanced decomposition), they may have been easier to detect by both groups, hence the lack of significant differences past the first week.

The results of Phase I demonstrated that observers differed from experts in their scoring of longitudinal decomposition, but it is unknown if the experts were influenced by any biases. Were the experts able to focus objectively on the data and ignore the contextual information of placement date? While it was suggested that a reason for the difference in scores at week one of Trial 2 was that experts were engaging in confirmation and anchoring biases because they may have already had expectations of what the donors would look like given their knowledge of placement time and decomposition progression (Bressan and Dal Martino 2005; Dror and Fraser-Mackenzie, 2009; Rossmo, 2009; Nakhaeizadeh, Dror et al., 2014; Nakhaeizadeh, Hanson et al., 2014), it may also be the case that due to their extensive experience working with human remains at the ARF, the experts were better able to reduce the effects of biasing information. They may have been focusing only on the decompositional characteristics and not the timing of placement. All of the experts previously collected daily data at the ARF and may have already noticed their tendencies to rely on certain types of information over others. Due to that, the experts may have developed strategies (e.g., focusing on goals to keep track of what information is relevant and irrelevant) that they could use to overcome the influence of the biasing information (Rossmo, 2009; Robbins and Judge, 2013). Essentially, the experts in this study might be cognitively blind to biasing placement information. However, Dror et al. (2006) discovered that only one in five (20%) experts in latent print examinations were able to ignore contextual information and that even with all of their training, contextual biases still revealed expert vulnerabilities. With that in mind, the experts in this study might have over or under-corrected their scores less than the observers because of their expert knowledge (Anderson, 2015). It is also possible that the bias of the experts simply takes a different, unknown form from that of the observers and this study was unable to capture it with the current methods. While this highlights concerns with expert bias that may lead to distorted scores, it does not imply that experts should not be used, just that their biases may be different from observers with less experience (Dror et al., 2006).

The over-and under-estimation of observers from Phase I become especially apparent when compared to the parallel results from the photograph assessment of Phase II, where no differences were detected between experts and observers in their Total Body Scores for the photos. The tasks in Phase I and Phase II were identical except that observers scored the photos only once in Phase II and were not given any information regarding the placement date or postmortem interval for any of the donors depicted in the photos. This means that, unlike Phase I, observers in Phase II were not given any biasing contextual information with which to influence their scores (Bressan and Dal Martino, 2005; Nakhaeizadeh, Dror et al., 2014; Nakhaeizadeh, Hanson et al., 2014). Thus, they would not feel the need to correct their scores and would not differ from the scores of experts. Furthermore, as the scoring system was originally developed using photographs, the photos may simply have better fit the parameters of the TBS categories than the field assessment (Megyesi, 2001; Dabbs et al., 2016).

These results have important implications for the use of TBS in longitudinal versus cross-sectional research. This dissertation has demonstrated that the Total Body Score method may not be appropriate for longer-term research applications because of the chance that contextual and confirmation biases may influence an examiners ability to provide accurate scores. This then impacts proper standard operating procedures at decomposition facilities, such as the ARF. Due to a variety of limitations, especially in staffing, it is sometimes impossible to be blinded to placement information (Kassin et al., 2013), particularly if the person in charge of donor placement is also the one collecting the daily data (as is common at decomposition facilities). A similar situation can occur in forensic casework where the examiners conducting the scene analysis can also examine evidence in the laboratory with all the prior knowledge from the scene, leading to potential biases during investigations (Dror, 2009; Kassin et al., 2013). Knowledge of the entire evidence collection and examination procedures can be critical for identification of remains, but decomposition scoring systems may need to be redesigned to limit misidentifications due to biases. This is the first large-scale project to compare evidence regarding the applicability of TBS to both field and photograph scenarios and to show that TBS may not be applicable to longitudinal field assessments.

Overall, Hypothesis 1, that the experts and observers will differ in their scoring of decomposition characteristics utilizing the Total Body Score method, is supported in both trials of Phase I. However, the pattern of that difference was not consistent and may have been based on the unique situations presented in each trial. Additionally, the biasing information unintentionally given to each observer that of the donors' time of placement, added to this inconsistency in directionality in Trials I and II.

The influence of cognitive bias on the scores of observers has been briefly discussed above in terms of their influence on accuracy. However, those results do not reflect the more direct measures of cognitive processes, specifically motivations and mood, which were utilized in this study. The results of Hypothesis 2 for Phase I suggest that perceived difficulty and pressure/tension differentially affected observers TBS decisions in both Trial 1 and Trial 2 with the additional significance of perceived competency in Trial 2. According to Ryan et al. (1990), the pressure/tension scale of the IMI is a negative predictor of intrinsic motivation. Intrinsic motivation is when an individual engages in a behavior out of enjoyment (Ryan et al., 1990). It is motivation that arises from within the individual. The pressure/tension scale included such items as “I felt tense while doing this activity,” “I was anxious while working on this task,” and “I felt relaxed” (Ryan et al., 1990). The significance of this is that observers reported being nervous while they scored, and since this is a negative predictor of intrinsic motivation, it indicates that they were nervous because of pressure to do well in an activity that they wanted to enjoy. This is similar to the findings of Charlton et al. (2010) who demonstrated that experienced fingerprint examiners reported heightened emotional states during their analyses, especially in high-profile cases where there is more pressure to obtain a match. This pressure to do well can bias results as an individual may feel internal pressure to get the answer that they perceive the supervisor desires (Charlton et al., 2010; Kassin et al., 2013). In the current study, where an observer was not asked to confirm a match or pattern like in fingerprint analysis, it is difficult for a specific answer to be desired as there are many possibilities in decomposition research. It can be interpreted then that observers wanted to get the ‘right’ answer which was the score obtained by either the majority of the observers in their scoring session group or the same score as Expert 1, who oversaw the sessions. On several occasions, at least one observer asked Expert 1 what score she obtained so she could compare her choice to that of Expert 1. Expert 1, in all cases refrained from telling the observer in order to prevent her from changing her score. Therefore, the pressure and anxiety recorded by observers in Trials 1 and 2 might stem from the desire to not be seen as incompetent or 'less than' in the eyes of Expert 1 on a task that, although they perceived as difficult, they nevertheless enjoyed.

The addition of perceived competency as a significant predictor occurred only in Trial 2. This variable measured the extent to which observers felt they did well at the scoring task. The differences between Trial 1 and Trial 2 that lead to competency being significant include 1) those individuals from Trial 1 who returned to participate in Trial 2 and already had experience scoring, therefore felt more competent in the task and/or 2) the longer duration of Trial 2 at ten weeks, versus seven weeks for Trial 1, provided observers with up to six additional scoring sessions. This extra experience may have led observers to feel more competent in their assessments. It is also possible that Trial 2 observers were confronted with less variability in decomposition characteristics for most of the trial, due to the colder temperatures, than they were in Trial 1 where changes occurred relatively rapidly.

The difference in motivations between photographs and field are most likely due to the differences inherent in the tasks. Full body photographs may have been seen as "easier" by observers because all of the information to score all segments is present in the photograph. Individuals may not have been as pressured or stressed to try to score with less information, resulting in a positive mood. However, partial photographs may have been more difficult because they are ambiguous and as such require additional decisions to be made such as 'Is there enough information to score any of the segments? If so, which ones?' Nakhaeizadeh, Dror et al.'s (2014) study of cognitive bias in trauma estimation found a greater variability in responses indicating trauma in ambiguous images (i.e., postmortem/taphonomic damage) than in images that depicted clear or no trauma. For these images, when provided with contextual information that the remains originated in a mass grave context, observers indicated the presence of more trauma than when they were told the remains originated at an archaeological site. The presence of ambiguity in decision-making has been shown to increase uncertainty and reliance on cognitive biases, especially confirmation bias (Dror et al., 2006; Nakhaeizadeh, Dror et al., 2014) and the current study supports this conclusion as negative affect was a significant predictor of score for the ambiguous, partial body photographs.

PHOTOGRAPHS VS. FIELD ASSESSMENTS

Field assessments are not the only way to assess decomposition. In many forensic cases, photographs may be the only evidence provided to an investigation, especially in cold cases where any soft tissue evidence will long be degraded or decomposed. As mentioned previously, experts and observers in this study did not significantly differ in terms of their photograph scores. Additionally, different predictors were significant in the photograph portion compared to field assessments. An unexpected result from Phase II, is that the level of observer experience was not influential in photo scoring. Nakhaeizadeh, Dror et al. (2014) showed a difference in trauma analysis between people with less than four years of experience (they indicated more trauma present) and those with five or more years of experience. However, contrary to Nakhaeizadeh, Dror et al.'s (2014) conclusions, scores in the current study, from observers who participated in Phase I did not differ from those observers who only participated in Phase II. This lack of significance actually demonstrates the versatility of this scoring system where individuals who have limited experience with this method and knowledge of human decomposition can still obtain reliable data from photographs. This is desirable since practitioners of this scoring system do not require, at least for forensic anthropology undergraduate students, years of field experience in human decomposition. Thus, a wide range of medicolegal, law enforcement, and academic professionals could conceivably use this system. However, TBS may be inappropriate to use in the field as it currently is constructed, due to ambiguity and because the system does not account for the relevant biases and subjectivities present during field assessment.

In terms of limited, cross-sectional applications, the TBS system appears to produce reliable information. As previously mentioned, Megyesi (2001) developed the method utilizing photographs from various forensic cases and therefore, the results from Phase II where experts and observers did not differ in their scores supports the utility of this method for that medium. This study also supports the findings of Dabbs et al. (2016) who described high levels of consistency in the scores of photographs from observers of varied experienced levels. Based on those results, they concluded that the TBS method could be eliminated as a potential source of error in estimating PMI from scores (Dabbs et al., 2016). While supporting the utility of TBS for photograph analysis, the current study highlights that this method is also being used for field-based longitudinal research and that results generated from those scenarios may be prone to bias and affecting TBS-derived PMI estimates. Forensic cases, however, represent a middle ground between the two scenarios tested in this dissertation: there is a field component, but it usually occurs over a very brief time period of time (e.g., a few days). Because they lack a truly longitudinal component, forensic cases may be able to utilize TBS reliably. The effect on the examiner would be similar to that of scoring a photograph because, if an anthropologist is called out to consult on a case, there typically may be very little contextual information that indicates an obvious PMI. Therefore, TBS would be appropriate for forensic casework. However, to be more confident in this conclusion, a future study should examine forensic cases or cross-sectional studies where TBS was utilized and where actual PMI was discovered to ensure those types of field-based assessments are accurate.

The final chapter (Chapter 7) will discuss avenues for future research.

CHAPTER SEVEN:

CONCLUSION

The application of the science of biological anthropology to forensic scenarios can aid in identifying remains and facilitate medicolegal investigations. In particular, forensic anthropologists strive to understand decomposition and its effect on remains in order to improve the chances of identification. By utilizing a well-known scoring system (Megyesi, 2001; Megyesi et al., 2005) and measures of motivation and mood (Ryan, 1982; McAuley et al., 1987; Watson, et al., 1988), this project has sought to understand how an individual's cognitive bias may affect their scoring accuracy.

Specifically, this dissertation has questioned whether observers' motivations and moods have a significant effect on TBS when compared to experts, implying that cognitive biases affect TBS accuracy. The data presented indicates that individual observer's motivations played a significant role in both field and photographic assessment as did contextual information regarding the placement time of the donors used for both Trial 1 and Trial 2; this information was not available in the photograph assessment. The lack of significant differences between observers and experts in Phase II may indicate that the TBS system is more appropriate for assessing photographs of human remains, rather than field applications.

CONTRIBUTION OF CURRENT RESEARCH TO FORENSIC ANTHROPOLOGY

As set out in Chapter 1, this dissertation sought to understand the factors that could lead to inaccurate estimations of decomposition and ultimately influence identification. The National Academy of Sciences (NAS) 2009 report encouraged forensic scientists to understand the role of cognitive biases in methodologies requiring human examiners. While forensic anthropology was not singled out within the report, cognitive biases could affect analyses of decomposition, which is one tool used to estimate time since death. This dissertation represents an extension of those NAS (2009) recommendations. As the results of any project are only as good as the available data, it is important to understand the quality of data that goes into the estimation of time since death from decomposition characteristics. Significantly, this project expanded on those analyses by, not just documenting the presence of bias, but reporting on the possible influential cognitive processes associated with this phenomenon. This project documented four main conclusions (Chapter 5):

1. Any information regarding time since death, or in this case, time since placement, can bias the analytical capabilities and conclusions of examiners.
2. Mood and motivation affected observer scores differently depending on the context. Mood did not explain scores in the longitudinal field scenarios, but it did impact scores derived from photographs. Similarly, motivations played a larger role in the longitudinal context than they did in the cross-sectional one.
3. The TBS scoring system when used on photographic evidence of human remains is less prone to bias and may be appropriate to be employed by a wide variety of medicolegal personnel.
4. The TBS scoring system may not sufficiently account for biases during longitudinal field assessments, despite their common use in such research at decomposition facilities. However, for cross-sectional studies, especially those involving

photographs, TBS appears to produce consistent assessments. Therefore, context of the scores should be considered when utilizing TBS in forensic cases.

It is possible that the analytical methods utilized in this dissertation might have been too restrictive or not restrictive enough to document differences in mood mentioned in conclusion two. However, if the scales or methods used to assess cognitive states were altered or broadened, then different results may be acquired in the future. Broadly speaking, when applying Total Body Score or any similar method, caution should be taken to limit the affect of contextual information, especially in a longitudinal setting.

FUTURE RESEARCH

This dissertation represents a starting point for a more complete evaluation of decomposition assessments in forensic anthropology. However, no research project is completely comprehensive and thus, several recommendations can be made for further research into this area.

As discussed in Chapter 6, observers' knowledge regarding placement times of the donors utilized in this study may have biased their decomposition assessments, leading them to differ from experts. This biasing effect was only observed during the Phase I longitudinal field trials and not during the photo assessment phase. To better understand the impact of placement information and if the magnitude is larger in longitudinal over cross-sectional studies, biasing contextual information can be given to observers during a photograph-scoring task, like the one used here. If the effects on scores are similar to the results observed in this dissertation, then prior contextual information in any scenario could be detrimental to an examiner's ability to provide accurate, reliable information. Nevertheless, safeguards and standard procedures need to be designed and employed to reduce the impact of biases in forensic anthropology research and casework. One suggestion to combat bias is to keep examiners blind to any contextual information, as is reasonably possible. For example, those who attend to a crime scene should not be the primary analysts in the laboratory due to the possibility that they may unknowingly retain biasing information from the scene (Kassin et al., 2013). While this may not be possible in all institutions due to staffing limitations, it is the ideal situation to ensure a reduced chance of bias.

In a similar vein, the development of additional measures to assess decomposition that focus on more objective standards can help to reduce biasing influences. These methods, such as those developed by Kenyhercz et al. (2017), focus more on the presence or absence of characteristics and require less interpretation by the observer. Furthermore, Kenyhercz et al.'s (2017) study was also developed in a longitudinal context where the observers are aware of the contextual placement information that biased observers in this study. Developing methods specifically in a longitudinal context and reducing the amount of subjectivity in the scoring system can reduce the impact of ambiguous situations and subsequently, lessen the influence of bias.

Furthermore, this study primarily examined undergraduate anthropology students who overwhelmingly had minimal exposure with human decomposition. While students of all experience levels are utilized in anthropological research, it may be more applicable to the forensic science community if more experienced observers were examined in the future. As discussed in this study, the experts did not complete the cognitive measures due to time constraints so their bias susceptibility was not fully measured. However, if a sample of individuals experienced in the assessment of human decomposition characteristics were to be examined, it may be discovered that they rely on other biases or other cognitive processes in their decision-making processes.

Lastly, it would be informative to investigate other psychological traits, such as specific personality traits, that might better predict scoring performance. Alternatively, it might be useful to try other measures of mood and motivation to assess whether they analyze those constructs better than the IMI and PANAS used in the project. It is possible that those measures did not fully encompass all of the influences on observers' scores and different measures may shed light on the mechanisms behind scoring decisions.

CONCLUDING REMARKS

The field of forensic anthropology is constantly and consistently expanding its methods of data collection and data analysis. The next decade of research into human decomposition will undoubtedly provide researchers with even more novel methodologies. Incorporating innovative methodologies and concepts from other fields of research such as psychology, and biology will further permit the field of biological anthropology to better understand human decomposition processes. While cognitive biases cannot be fully eliminated, they can be studied, measured, and accounted for. Furthermore, extensive training and experience might play a vital role in minimizing biasing effects, the ability to recognize possible external influences, and strategies for limiting their effects must also be included in education and training of forensic anthropologists.

LIST OF REFERENCES

Aarts H, Dijksterhuis A. Category activation effects in judgment and behavior: The moderating role of perceived comparability. Br J Soc Psychol 2002; 41:123-138.

Adlam RE, Simmons T. The effect of repeated physical disturbance on soft tissue decomposition—are taphonomic studies an accurate reflection of decomposition? J Forensic Sci 2007; 52:1007-1014.

Anderson GS. Minimum and maximum development rates of some forensically important Calliphoridae (Diptera). J Forensic Sci 2000; 45:824-832.

Anderson JR. Expertise. In: Anderson JR, editor. Cognitive psychology and its implications. 8th rev. ed. New York, NY: Worth Publishers, 2015; 210-236.

Arnold J, Silvester J, Randall R, Patterson F, Robertson TI, Buphres B, et al. Work psychology: understanding human behavior in the workplace. 4th rev. ed. New York, NY: Prentice Hall, 2005.

Asch SE. Forming impressions of personality. J Abnorm Soc Psych 1946; 41:258-290.

Ask K, Granhog PA. Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. J Invest Psychol Off 2005; 2:43-63.

Aturaliya S, Lukasewycz A. Experimental forensic and bioanthropological aspects of soft tissue taphonomy: 1. Factors influencing postmortem tissue dessication rate. J Forensic Sci 1999; 44:893-896.

Augier M. Simon says: Bounded rationality matters. J Manag Inq 2001; 10:268-275.

Baden M, Hennessee J. Unnatural death: confessions of a medical examiner. New York, NY: Ballantine Books, 1989.

Balcetis E, Dunning D. See what you want to see: Motivational influences on visual perception. J Pers Soc Psychol 2006; 91:612-625.

Bass WM. Outdoor decomposition rates in Tennessee. In Haglund WD, Sorg MH, editors. Forensic taphonomy: The postmortem fate of human remains. New York, NY: CRC Press, 1997; 181-186.

Bazerman MH, Moore DA. Judgment in managerial decision making. 7th rev. ed. Hobokon, NJ: Wiley, 2008.

Beasley JC, DeVault TL, Retamosa MI, Rhodes Jr. OE. A hierarchical analysis of habitat selection by raccoons in northern Indiana. J Wildl Manag, 2007a; 71:1125-1133.

Beasley JC, DeVault TL, Rhodes Jr. OE. Home-range attributes of raccoons in a fragmented agricultural region of northern Indiana. The J Wildl Manag 2007b; 71:844-850.

Berryman HE. Disarticulation pattern and tooth mark artifacts associated with pig scavenging of human remains: a case study. In: Haglund WD, Sorg MH, editors. Advances in forensic taphonomy: Method, theory, and archaeological perspectives. Boca Raton, FL: CRC Press, 2002; 487-495.

Bieber P. Fire Investigation and Cognitive Bias. In: Jamieson A, Moenssens AA, editors. Wiley encyclopedia of forensic science Vol 1. Hoboken, NJ: John Wiley & Sons, Ltd, 2014;1-8.

Boedecker P. Hierarchical linear modeling with maximum likelihood, restricted maximum likelihood, and fully Bayesian estimation. Pract Assess Res Eval 2017; 22:1-19.

Brandimonte MA, Bruno N, Collina S. Cognition. In: Pawlik P, D'Ydewalle G, editors. Psychological concepts: An international historical perspective. Hove, UK: Psychology Press, 2006; 11-27.

Bressan P, DalMartello MF. Talis pacter, talis filius: Perceived resemblance and the belief in genetic relatedness. Psychol Sci 2002; 13: 213-218.

Brown PD. Some hunches about intuition. New York Times 2007 Nov 17; Sect. B:5.

Bytheway JA, Miller N, Steadman DW, Sauerwein K, Wescott D, Clemmons CM, et al. Validation of TBS/ADD equation at 100, 300, 500, and 1000 ADD on 30 human subjects with known PMI from three human decompositional facilities. Forensic Anthropol. In press.

Campobasso CP, DiVella G, Introna F. Factors affecting decomposition and Diptera colonization. Forensic Sci Int 2001; 120:18-27.

Catanese G, Bloom T. Recovery of a mummified pregnant woman from a 55 gallon drum more than 30 years after her death. Am J Forensic Med Pathol 2002; 23:245-247.

Charlton D, Dror IE, Fraser-Mackenzie PAF. Emotional experiences and motivating factors associated with fingerprint analysis. J Forensic Sci 2010; 55:385-393.

Charman SD, Gregory AH, Carlucci M. Exploring the diagnostic utility of facial composites: Beliefs of guilt can bias perceived similarity between composite and suspect. J Exp Psychol Appl 2009; 15:76-90.

Christensen AM, Passalacqua NV, Bartelink EJ. Forensic anthropology: Current methods and practice. San Diego, CA: Elsevier Academic Press, 2014

Clark MA, Wornell MR, Pless JE. Postmortem changes in soft tissues. In: Haglund WD, Sorg MH, editors. Forensic taphonomy: Postmortem fate of human remains. New York, NY: CRC Press, 1997; 151-164.

Coe JI. Postmortem chemistry update: emphasis on forensic application. Am J Forensic Med Pathol 1993; 14:92-93.

Corbeil RR, Searle SR. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. Technometrics 1976; 18:31-38.

Coutts LM, Gruman JA. Applying social psychology to organizations. In: Schneider FW, Gruman JA, Coutts LM, editors. Applied social psychology: Understanding and addressing social and practical problems. 2nd rev. ed. Los Angeles, CA: Sage, 2012; 217-244.

Crawford JR, Henry JD. The positive and negative affect schedule (PANAS): Construct validity, measurement properties, and normative data in a large non-clinical sample. Br J Clin Psychol 2004; 43:245-265.

Dabbs GR, Connor M, Bytheway JA. Interobserver reliability of the Total Body Score system for quantifying human decomposition. J Forensic Sci 2016; 61:445-451.

Dautartas AM. The Effect of Various Coverings on the Rate of Human Decomposition [thesis]. Knoxville, TN: University of Tennessee, 2009.

Dautartas A, Kenyhercz MW, Vidoli GM, Jantz LM, Mundorff A, Steadman DW. Differential decomposition among pig, rabbit, and human remains. J Forensic Sci 2018: 1-11.

Dawes J. Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point, and 10-point scales. Int J Mark Res 2008; 50:61-77.

David D, Midea M, Opre A. The information-processing approach to the human mind: Basics and beyond. J Clin Psychol 2004; 60:353-388.

Dayananda R, Kiran J. Entomotoxicology. Int J Med Toxicol Forensic Med. 2013; 3:71-74.

Deci EL, Eghrari H, Patrick BC, Leone D. Facilitating internalization: The self-determination theory perspective. J Pers 1994; 62:119-142.

DiVellis RF. Scale development: theory and applications. Los Angeles, CA: Sage, 2012; 109-110.

DiMaio D, DiMaio VJM. Forensic Pathology. 2nd rev. ed. New York, NY: CRC Press, 2001.

Dror IE. How can Francis Bacon help forensic science? The four idols of human biases. Jurimetrics J. 2009; 50:76-90.

Dror IE. A hierarchy of expert performance. J Appl Res Mem Cogn 2016; 5:121-127.

Dror IE, Charlton D. Why experts make errors? J Forensic Identif 2006; 56:600-616.

Dror IE, Cole SA. The vision in blind justice: Expert perception, judgment, and visual cognition in forensic pattern recognition. Psychon Bull Rev 2010; 17:161-167.

Dror IE, Fraser-Mackenzie P. Cognitive biases in human perception, judgment, and decision-making: Bridging theory and the real world. In: Rossmo DK, editor. Criminal Investigative Failures. Boca Raton, FL: CRC Press, 2009; 53-67.

Dror IE and Hampikian G. Subjectivity and bias in forensic DNA mixture interpretation. Sci Justice 2011; 51:204-208.

Dror IE, Rosenthal R. Meta-analytically quantifying the reliability and bias ability of forensic experts. J Forensic Sci 2008; 53:900-903.

Dror IE, Charlton D, Peron AE. Contextual information reveals experts vulnerable to making erroneous identification. Forensic Sci Int. 2006; 156:74-78.

Dror IE, Peron A, Hind S, Charlton D. When emotions get the better of us: The effect of contextual top-down processing on matching fingerprints. Appl Cogn Psychol 2005: 19:799-809.

Efremov JA. Taphonomy: New branch of paleontology. Pan-Amer Geol 1940; 74:81-93.

Evans WED. The chemistry of death. Springfield, IL: Charles C. Thomas Publisher, 1963.

Evans JBT, Barston JL, Pollard P. On the conflict between logic and belief in syllogistic reasoning. Mem Cogn 1983; 11:295-306.

Festinger L. A theory of social comparison processes. Hum Relat 1954; 7:117-140.

Fielder S, Graw M. Decomposition of buried corpses with special reference to the formation of adipocere. Naturwissenschaften 2003; 90:291: 300.

Fiske S, Taylor SH. Social Cognition, 2nd rev. ed. New York, NY: McGraw-Hill, 1991.

Fischhoff B, Slovic P, Lichtenstein S. Knowing with certainity: The appropriateness of extreme confidence. J Exp Psychol Hum Percept Perform 1977; 3:552-564.

Forbes SL, Stuart BM, Dadour IR, Dent BB. A preliminary investigation of the stages of adipocere formation. J Forensic Sci 2004; 49:1-9.

Found B, Ganas J. The management of domain irrelevant context information in forensic handwriting examination casework. Sci Justice 2013; 53:154-158.

Freund T, Kruglanski AW, Shpitzajzen A. The freezing and unfreezing of impressional primacy: Effects of the need for structure and the fear of invalidity. Pers Soc Psychol Bull 1985; 11:479-487.

Galloway A, Birkby W, Jones AM, Henry TE, Parks BO. Decay rates of human remains in an arid environment. J Forensic Sci 1989; 34:607-616.

Garrett BL. Convicting the innocent: Where criminal prosecutions go wrong. Cambridge, MA: Harvard University Press, 2011.

Giannelli PC. Wrongful convictions and forensic science: The need to regulate crime labs. N C L Rev 2007; 86:163-235.

Gigerenzer G. Why heuristics work. Perspect Psychol Sci 2008; 3:20-29.

Gill-King H. Chemical and Ultrastructural Aspects of Decomposition. In: Haglund WD, Sorg MH, editors. Forensic taphonomy: Postmortem fate of human remains. New York, NY: CRC Press, 1997; 93-108

Goethals GR, Messick DM, Allison ST. The uniqueness bias: Studies of constructive social comparison. In: Souls J, Will TA, editors. Social Comparison and Contemporary Theory and Research. Hillsdale, NJ: Erlbaum, 1991.

Gonzales TA, Vance M, Helpern M, Umberger CJ. Legal medicine, Pathology and Toxicology. New York: Appleton-Century Crofts Inc, 1954.

Haber L, Haber RN. Scientific validation of fingerprint evidence under Daubert. Law Prob Risk 2008; 7:87-109.

Haglund WD. Dogs and Coyotes: postmortem involvement with human remains. In: Haglund WD, Sorg MH, editors. Forensic taphonomy: Postmortem fate of human remains. New York, NY: CRC Press, 1997a; 367-382.

Haglund WD. Rodents and human remains. In: Haglund WD, Sorg MH, editors. Forensic taphonomy: Postmortem fate of human remains. New York, NY: CRC Press, 1997b; 405-414.

Haglund WD, Reay DT, Swindler DR. Tooth mark artifacts and survival of bones in animal scavenged human skeletons. J Forensic Sci 1988; 33:985-997.

Haglund W, Reay DT, Swindler DR. Canid scavenging/disarticulation sequence of human remains in the Pacific Northwest. J Forensic Sci 1989; 34:587-606.

Haglund WD, Sorg MH. Method and theory of forensic taphonomic research. In: Haglund WD and Sorg MH, editors In: Haglund WD, Sorg MH, editors. Forensic taphonomy: Postmortem fate of human remains. New York, NY: CRC Press, 1997; 13-26.

Halfpenny JC, Bruchac J. Scats and tracks of the southeast: A field guide to the signs of seventy wildlife species. Guilford, CT: The Globe Pequot Press, 2002; 94-95.

Harle L. Postmortem changes. 2012. http://www.pathologyoutlines.com/topic/forensicspostmortem.html (accessed 20 May 2016).

Hart W, Albarracin D, Eagly AH, Brechan I, Lindberg M, Merrill L. Feeling validated versus being correct: A meta-analysis of selective exposure to information. Psychol Bull. 2009; 135:555-588.

Haynes, G. Utilization and skeletal disturbances of North American prey carcasses. Arctic 1982; 35:266-281.

Hefner JT, Ousley SD. Statistical Classification Methods for Estimating Ancestry using Morphoscopic Traits. J Forensic Sci 2014; 59:883-890.

Heiberger RM, Robbins NB. Design of diverging stacked bar charts for Likert scales and other applications. J Stat Softw 2014; 57:1-32.

Heuer Jr RJ. Psychology of Intelligence Analysis. Washington DC: Center for Study of Intelligence, Central Intelligence Agency, 1999.

Hicks JA, Cicero DC, Trent J, Burton CM, King LA. Positive affect, intuition, and feelings of meaning. J Pers Soc Psychol 2010; 98:967-979.

IBM Corp. IBM SPSS for Macintosh Version 23.0. Armonk, NY: IBM Corporation, 2016.

Janaway RC. The decay of buried human remains and their associated materials. In: Hunter J, Roberts C, Martin A, editors. Studies in Crime: An Introduction to Forensic Archaeology. London: Batsford, 1996; 58-85.

Janaway RC, Percival SL, Wilson AS. Decomposition of human remains. In: Percival SL, editor. Microbiology and aging: clinical manifestations. Flintshire, UK: Springer, 2009; 313-334.

Jonas E, Schultz-Hardt S, Frey D, Thelen N. Confirmation bias in sequential information search after preliminary decisions. J Pers Soc Psychol 2001; 80:557-571.

Kahneman D. A perspective on judgment and choice: Mapping bounded rationality. Am Psychol 2003a; 58:697-720

Kahneman D. Maps of bounded rationality: Psychology for behavioral economics. Am Econ Rev 2003b; 93:1449-1475.

Kassin SM, Dror IE, Kukucka J. The forensic confirmation bias: Problems, perspectives, and proposed solutions. J Appl Res Mem Cogn 2013; 2:42-52.

Kenyhercz MW, Steadman DW, Gundel AC. A new approach to estimating Accumulated Degree Days (ADD): Using binary observations from various regions of the body with random forest modeling (RFM) and geographic information systems (GIS). In: Proceedings of the 69th American Academy of Forensic Sciences Annual Meeting, 2017 February13-18; New Orleans, LA; 246-247.

Kelley HH. Attribution theory in social psychology. In: Levine D, editor. Nebraska Symposium on motivation. Lincoln, NE: University of Nebraska Press, 1967.

Kelley HH. Causal schemata and the attribution process. New York, NY: General Learning Press, 1972.

Kerstholt J, Eikelboom A, Dijkman T, Stoel R, Hermsen R, van Leuven B. Does suggestive information cause a confirmation bias in bullet comparisons? Forensic Sci Int 2010; 198: 138-142.

Kline P. The handbook of psychological testing. 2nd rev. ed. London, UK: Routledge, 2000; 13.

Klippel WE, Synstelien JA. Rodents as taphonomic agents: bone gnawing by brown rats and gray squirrels. J Forensic Sci 2007; 52:765-773.

Komar D, Beattie O. Identifying bird scavenging in fleshed and dry remains. Can Soc Forensic Sci J 1998; 31:177-188.

Krane DE, Ford S, Gilder JR, Inman K, Jamieson A, Koppl R et al. Sequential unmasking: A means of minimizing observer effects in forensic DNA interpretation. J Forensic Sci 2008; 53:1006-1007

Krugar J, Dunning D. Unskilled and aware of it: How difficulties in recognizing one's own incompetence leads to inflated self-assessments. J Pers Soc Psychol 1999; 77:1121-1134.

Kruglanski AW, Freund T. The freezing and unfreezing of lay inferences: Effects on impressional primacy, ethnic stereotyping, and numerical anchoring. J Exp Soc Psychol 1983; 19:448-468.

Kruglanski AW, Gigerenzer G. Intuitive and deliberate judgments are based on common principles. Psychol Rev 2011; 118:97-109.

Kunda Z. The case for motivated reasoning. Psychol Bull 1990; 108: 480-498.

Larrick RP, Bunson KA, Soll JB. Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not). Organ Behav Hum Decis Process 2007; 102:76-94.

Lord WD. Case histories of the use of insects in investigations. In: Catts PE, Haskell NH, editors. Entomology and death: A procedural guide. Clemson, SC: Joyce's Print Shop, Inc, 1990.

Love JC, Marks MK. Taphonomy and time: Estimating the postmortem interval. In: Steadman DW, editor. Hard evidence: Case studies in forensic anthropology. Upper Saddle River, NJ: Pearson Education Inc, 2003; 160-175.

Macchiarelli L, Feola T. Medicina legale. Torino, ITL: Minerva Medica, 1995.

Mann RW, Bass WM, Meadows L. Time since death and decomposition of the human body: variables and observations in case and experimental field studies. J Forensic Sci 1990; 35:103-111.

Mant AK. Knowledge acquired from post-war exhumations. In: Boddington A, Garland AN, Janaway RC, editors. Death, decay and reconstruction: Approaches to archaeology and forensic sciences. Manchester, England: Manchester University Press, 1987; 65-80.

Marks MK, Love JC, Dadour IR. Taphonomy and time: Estimating the postmortem interval. In: Steadman DW, editor. Hard evidence: Case studies in forensic anthropology. Upper Saddle River, NJ: Pearson Education, 2009; 167-180.

Martinko MJ, Harvey P, Dasborough MT. Attribution theory in the organizational sciences: A case of unrealized potential. J Organ Behav 2011; 32:144-149.

Mayer RG. 2012 Embalming: history, theory, and practice. Fifth ed. The McGraw-Hill Companies, Inc. New York, NY.

McAuley E, Duncan T, Tammen VV. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. Res Q Exerc Sport 1989; 60:48-58.

Megyesi MS. The effects of temperature on the decomposition rate of human remains [thesis]. Indianapolis, IN: University of Indianapolis, 2001.

Megyesi MS, Nawrocki SP, Haskell NH. 2005. Using accumulated degree-days to estimate the postmortem interval from decomposed human remains. J Forensic Sci 2005; 50:1-9.

Micozzi MS. Postmortem change in human and animal remains. Springfield IL: Charles C. Thomas Publishing, 1991.

Micozzi MS. Frozen environments and soft tissue preservation. In: Haglund WD, Sorg M., editors. Forensic taphonomy: The postmortem fate of human remains. New York, NY: CRC Press, 1997; 171-180.

Miller, RA. The effects of clothing on human decomposition: implications for estimating time since death [thesis]. Knoxville, TN: University of Tennessee Knoxville, 2002.

Mnookin JL. The validity of latent fingerprint identification: Confessions of a fingerprinting moderate. Law Probab Risk 2008; 7:127-141.

Myburg J, L'abbe EN, Steyn M, Becker PJ. Estimating the postmortem interval (PMI) using accumulated degree-days (ADD) in a temperate region of South Africa. Forensic Sci Int 2013; 229:165.e1-165.e6.

Myers D. Intuition: Its powers and perils. New Haven, CT: Yale University Press, 2002.

Nakhaeizadeh S, Dror IE, Morgan RM. Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias. Sci Jus 2014a; 54:208-214.

Nakhaeizadeh S, Hanson I, Dozzi N. The power of contextual effects in forensic anthropology: A study of biasability in the visual interpretations of trauma analysis in skeletal remains. J Forensic Sci 2014b; 59:1177-1183

National Academy of Science. *Strengthening Forensic Science in the United States: A Path Forward.* Washington DC: National Academies Press, 2009.

Nawrocka M, Fratczak K, Matuszewski S. Inter-rater reliability of total body score – A scale for quantification of corpse decomposition. J Forensic Sci 2016; 61:798-802.

Neisser U. Cognition and Reality. San Francisco, CA: Freeman, 1976.

Nisbett R and Wilson T. Telling more than we can know: Verbal reports on mental processes. Psychol Rev 1977; 84:231-259.

Page M, Taylor J, Blenkin M. Context effects and observer bias – implications for forensic odontology. J Forensic Sci 2012; 57:108-112.

Palmbach TM, Hammond DL, Rennison A, Ross A. Reinforcing, challenging, and transforming our forensic science thinking: Progress and constraints five years after the NAS report. In: Proceedings of the 66th American Academy of Forensic Sciences Annual Meeting, 2014 February 17-22; Seattle, WA.

Perper JA. Time of death and changes after death, Part I: Anatomical considerations. In: Spitz WU, editor. Spitz and Fisher's Medicolegal investigation of death: Guidelines for the application of pathology to crime investigation. Springfield, IL: Charles C. Thomas, 1993; 87-127

Plant RW, Ryan RM. Intrinsic motivation and the effects of self-consciousness, self-awareness, and ego-involvement: An investigation of internally-controlling styles. J Pers 1985; 53:435-449.

Plous S. The psychology of judgment and decision making. New York, NY: McGraw-Hill, 1993; 217.

Polson CJ. The essentials of forensic medicine. London, UK: English Universities Press Limited, 1996.

Reeves NM. Taphonomic effects of vulture scavenging. J Forensic Sci 2009; 4:523- 528.
Ricketts DR. Scavenging effects and scattering patterns on porcine carcasses in Eastern Massachusetts [thesis]. Boston, MA: Boston University School of Medicine, 2013.

Risinger DM, Saks MJ, Thompson WC, Rosenthal R. The Daubert/Kumho implications of observer effects in forensic sciences: Hidden problems of expectation and suggestion. Cal L Rev 2002; 90:3-56.

Robbins SP, Judge TA. Perception and individual decision making. In: Robbins SP, Judge TA, editors. Organizational Behaviour. Boston, MA: Pearson, 2013; 165-199.

Robbins SP, Langton N. Organizational Behavior. Upper Saddle, NJ: Prentice Hall, 2001.

Rodriguez W, Bass WM. Insect activity and its relationship to decay rates of human cadavers in east Tennessee. J Forensic Sci 1983; 28:423-432.

Rossmo DK. Cognitive biases: Perception, Intuition, and tunnel vision. In: Rossmo DK, editor. Criminal Investigative Failures. Boca Raton, FL: CRC Press, 2009.

Roussere GP, Murray WJ, Raudenbush CB, Kutilek MJ, Levee DJ, Kazacos KR. Raccoon roundworm eggs near homes and risk for Larva Migrans disease in California communities. Emerg Infect Diseases. 2003; 9:1516- 1522.

Russo JE, Carlson KA, Meloy MG. Choosing an inferior alternative. Psychol Sci 2006; 17:899-904.

Ryan RM. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. J Pers Soc Psychol 1982; 43:450-461.

Ryan RM, Connell JP, Plant RW. Emotions in non-directed text learning. Learn Indiv Differ 1990; 2:1-17.

Ryan RM, Koestner R, Deci EL. Varied forms of persistence: When free-choice behavior is not intrinsically motivated. Motiv Emot 1991; 15:185-205.

Ryan RM, Mims V, Koestner R. Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. J Pers Soc Psychol 1983; 45:736-750.

Saks MJ, Koehler JJ. The coming paradigm shift in forensic identification science. Science 2005; 309:892-895.

Schiffer B, Champod C. The potential (negative) influence of observational biases at the analysis stage of fingermark individualisation. Forensic Sci Int 2007; 167:116-120.

Shafir E, LeBoeuf RA. Rationality. Annu Rev Psychol 2002; 53:491-517.

Shah AK, Oppenheimer DM. Heuristics made easy: An effort reduction framework. Psychol Bull 2008; 134:207-222.

Simmons JP, LeBoeuf RA, Nelson LD. The effect of accuracy motivation on anchoring and adjustment: Do people adjust from their provided anchors? J Pers Soc Psychol 2010; 99:917-932.

Simon D, Thompson WC, Kassin S, Sulner A, Scheck B. Human factors in forensic science: Why cognitive bias can lead to flawed expert opinions and testimony, how its influence can be minimized, and what challenges testifying experts and judges can expect to face. Proceedings of the 67th Plenary American Academy of Forensic Sciences Annual Meeting, 2015 February 18-21; Orlando, FL.

Simons DJ, Chabris CF. Gorillas in our midst: Sustained inattential blindness for dynamic events. Perception 1999; 28:1059-1074.

Sledzik P. Forensic taphonomy: Postmortem decomposition and decay. In: Reichs KJ, editor. Forensic osteology: Advances in the identification of human remains. Springfield IL: Charles C. Tomas, 1998; 109-119.

Smith JK. Raccoon scavenging and the taphonomic effects on early human decomposition and PMI estimation [thesis]. Knoxville, TN: University of Tennessee, 2015.

Smith, SA. Studies in identification, no. 3. In: Rathburn, TA, Buikstra JE, editors. Human identification: Case studies in forensic anthropology. Springfield, Ill: Charles C. Thomas Publishing, 1984; 19-27.

Soon CS, Brass M, Heinze HJ, Hayes JD. Unconscious determinants of free decisions in the human brain. Nat Neurosci 2008; 11:543-545.

Souza MJ, Ramsay EC, Patton S, New JC. Baylisascaris procyonis in raccoons (Procyon lotor) in eastern Tennessee. J Wildl Dis 2009; 45:1231-1234.

Spitz WU, Fischer RS. Spitz and Fisher's Medicolegal investigation of death: Guidelines for the application of pathology to crime investigation. Springfield, IL: Charles C. Thomas, 1993.

Srnka CF. The effects of sun and shade on the early stages of human decomposition [thesis]. Knoxville, TN: University of Tennessee Knoxville, 2003.

Steadman DW, Worne H. Case report: Canine scavenging of human remains in an indoor setting. Forensic Sci Int 2007; 173:78-82.

Stelfox P, Pease K. Cognition and detection: Reluctant bedfellows? In: Smith MJ, Tilley N, editors. Crime science: New approaches to preventing and detecting crime. Cullompton, Devon, UK: Willan Publishing, 2005; 191-207.

Stewart TA. How to think with your gut. *Business 2.0* 2002; 1-6.
Stuart J. The effect of human body mass on the rate of decomposition. In: Proceedings of the 55th American Academy of Forensic Sciences Annual Meeting, 2003 February 17-22; Chicago, IL; 257.

Suckling JK. A longitudinal study on the outdoor human decomposition sequence in Central Texas [thesis]. San Marcos, TX: Texas State University, 2011.

Suckling JK, Spradley MK, Godde K. A longitudinal study on human outdoor decomposition in Central Texas. J Forensic Sci 2015; 61:19-25.

Svenson O. Are we less risky and more skillful than our fellow drivers? Acta Psychol 1981; 47:143-148.

Synstelien JA. Studies in taphonomy: bone and soft tissue modifications by postmortem scavengers [Dissertation]. Knoxville, TN: University of Tennessee, 2015.

Tabachnick BG, Fidell LS. Using multivariate statistics 6th rev. ed. Boston, MA: Pearson, 2013.

Taylor SE Brown J. Illusion and well-being: A social psychological perspective on mental health. Psychol Bull 1988; 103:193-210.

Tellegan A. Structures of mood and personality and their relevance to assessing anxiety with emphasis on self-report. In: Tuma AH, Maser JD, editors. Anxiety and the anxiety disorders. Hillsdale, NJ: Erlbaum and Associates., 1985; 681-706.

Toure-Tillery M, Fishbach A. How to measure motivation: A guide for the experimental social psychologist. Soc Pers Psychol Compass. 2014; 8:328-341.

Ubelaker DH, Zarenko KM. Adipocere: what is known after two centuries of research. Forensic Sci Int 2011; 208:167-172.

Ulery BT, Hicklin RA, Buscaglia J, Roberts MA. Repeatability and reproducibility of decisions made by valent fingerprint examiners. PlosOne 2012; 7:1-12.

Van Alphren A, Halfens R, Hasman A, Imbos T. Likert or Rasch? Nothing is more applicable than good theory. J Adv Nurs. 1994; 20:196-201.

Vass AA, Bass WM, Wolt JD, Foss JE, Ammons JT. Time since death determinations of human cadavers using soil solution. J Forensic Sci 1992; 57:1236-1253.

Vass AA, Smith RR, Thompson CV, Burnett MN, Dulgerian N, Eckenrode BA. Odor analysis of decomposing buried human remains. J Forensic Sci 2008; 53:384-391.

Watson D, Clark LA. Negative affectivity: The disposition to experience aversive emotional states. Psychol Bull 1984; 96:465-490.
Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. J Pers Soc Psychol 1988; 54:1063-1070.

Weigand H. Rebuilding a life: The wrongfully convicted and exonerated. Pub Int L J 2009; 18:427-437.

Westfall P, Henning KSS. Understanding advanced statistical methods. Boca Raton, FL: Chapman and Hall-CRC Press, 2013.

Wilder HH. Notes on the Indians of southern Massachusetts. Am Anthropol 1923; 25:197-218.

Willey P, Snyder L. Canid modification of human remains: Implications for time since-death estimations. J Forensic Sci 1989; 34:894-901.

Wilson-Taylor RJ. Time since death estimation and bone weathering: the postmortem interval. In: Tersigni-Tarrant MTA, Shirley NR, editors. Forensic anthropology: An introduction. Boca Raton, FL: CRC Press, 2013; 339-380.

Zhou C, Byard RW. Factors and processes causing accelerated decomposition in human cadavers – an overview. J Forensic Leg Med 2011; 18:6-9.

APPENDICES

APPENDIX A:

SCORING MATERIALS

ID #:
Date:

Scoring data collection sheet

Donation	Head & Neck	Trunk	Limbs	TBS	Comments

Notes:

Categories and Corresponding Descriptions for Scoring the Head & Neck	**Assigned Score**
A. Fresh	
1. Fresh, no discoloration.	1 pt
B. Early Decomposition	
1. Pink-White appearance with skin slippage and some hair loss.	2 pts
2. Gray to green discoloration: some flesh still relatively fresh.	3 pts
3. Discoloration and /or brownish shades particularly at edges, drying of nose, ears, and lips.	4 pts
4. Purging of decompositional fluids out of eyes, ears, nose, mouth, some bloating of neck and face may be present.	5 pts
5. Brown to black discoloration of flesh.	6 pts
C. Advanced decomposition	
1. Caving in of the flesh and tissues of eyes and throat.	7 pts
2. Moist decomposition with bone exposure less than one half that of the area being scored.	8 pts
3. Mummification with bone exposure less than one half that of the area being scored.	9 pts
D. Skeletonization	
1. Bone exposure of more than half of the area being scored with greasy substances and decomposed tissue.	10 pts
2. Bone exposure of more than half of the area being scored with desiccated or mummified tissue.	11 pts
3. Bones largely dry, but retaining some grease.	12 pts
4. Dry bone.	13 pts

Categories and Corresponding Descriptions for Scoring the Trunk	**Assigned Score**
A. Fresh	
1. Fresh, no discoloration.	1 pt
B. Early Decomposition	
1. Pink-white appearance with skin slippage and marbling present.	2 pts
2. Gray to green discoloration: some flesh still relatively fresh.	3 pts
3. Bloating with green discoloration and purging of decompositional fluids.	4 pts
4. Postbloating following release of the abdominal gases, with discoloration changing from green to black.	5 pts
C. Advanced Decomposition	
1. Decomposition of tissue producing sagging of flesh; caving in of the abdominal cavity	6 pts
2. Moist decomposition with bone exposure less than one half that of the area being scored.	7 pts
3. Mummification with bone exposure less than one half that of the area being scored.	8 pts
D. Skeletonization	
1. Bones with decomposed tissue, sometimes with body fluids and grease still present.	9 pts
2. Bones with desiccated or mummified tissue covering less than one half of the area being scored.	10 pts
3. Bones largely dry, but retaining some grease.	11 pts
4. Dry bone.	12 pts

Categories and Corresponding Descriptions for Scoring the Limbs	**Assigned Score**
A. Fresh	
1. Fresh, no discoloration.	1 pt
B. Early Decomposition	
1. Pink-white appearance with skin slippage and marbling present.	2 pts
2. Gray to green discoloration: some flesh still relatively fresh.	3 pts
3. Bloating with green discoloration and purging of decompositional fluids.	4 pts
4. Postbloating following release of the abdominal gases, with discoloration changing from green to black.	5 pts
C. Advanced Decomposition	
1. Decomposition of tissue producing sagging of flesh; caving in of the abdominal cavity	6 pts
2. Moist decomposition with bone exposure less than one half that of the area being scored.	7 pts
3. Mummification with bone exposure less than one half that of the area being scored.	8 pts
D. Skeletonization	
1. Bones with decomposed tissue, sometimes with body fluids and grease still present.	9 pts
2. Bones with desiccated or mummified tissue covering less than one half of the area being scored.	10 pts
3. Bones largely dry, but retaining some grease.	11 pts
4. Dry bone.	12 pts

ID #:
Date:

Perceptions of Decomposition Scoring Methods in Forensic Anthropology
Post-Scoring Survey

1. What characteristics did you use to help you choose your scores?

2. What decomposition characteristics do you think are the most important for anthropologists to consider?

3. Additional Comments:

4. For each of the following statements, please indicate how true it is for you, using the following scale:

1	2	3	4	5	6	7
not at all true			somewhat true			very true

______I enjoyed doing this activity very much

______I think I am pretty good at this activity.

______I put a lot of effort into this.

______I did not feel nervous at all while doing this.

______After working at this activity for awhile, I felt pretty competent.

______This activity was fun to do.

______I felt pressured while doing these.

______I am satisfied with my performance at this task.

______I didn't try very hard to do well at this activity.

______I thought this was a boring activity.

______I was very relaxed in doing these.

______I was pretty skilled at this activity.

______While I was doing this activity, I was thinking about how much I enjoyed it.

______It was important to me to do well at this task.

______I was anxious while working on this task.

______I felt this task got more difficult over time.

______I think I did pretty well at this activity, compared to other students.

______I didn't put much energy into this.

______This activity did not hold my attention at all.

______I felt very tense while doing this activity.

______This task was very tedious

______I tried very hard on this activity.

______I would describe this activity as very interesting.

______This was an activity that I couldn't do very well.

______I thought this activity was quite enjoyable.

______This task was easy to do.

5. This scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you feel this way right now, that is, at the present moment. Use the following scale to record your answers:

1	2	3	4	5
very slightly or not at all	a little	moderately	quite a bit	extremely

_____interested	_____irritable
_____distressed	_____alert
_____excited	_____ashamed
_____upset	_____inspired
_____strong	_____nervous
_____guilty	_____determined
_____scared	_____attentive
_____hostile	_____jittery
_____enthusiastic	_____active
_____proud	_____afraid

Demographics

6. Age: Sex:

7. UT Affiliation (circle one): Student Faculty Staff

8. If ‘student’ select your level: Freshman Sophomore Junior Senior Graduate Student
Other____________________________________

9. FAC involvement (circle all that apply):
Processor ARF recoveries Research assistant Intake Placements Bass Collection
Other:____

10. How long have you worked with the FAC? ______________________________

APPENDIX B:

SAS HIERARCHICAL LINEAR MODELS CODE

The following SAS v. 23.0 (IBM, Armonk, NY, 2016) code was used to construct the Hierarchical Linear Models used in this dissertation. Each section of code is labeled with the specific phase/trial and table to which it refers.

Phase I: Trial 1 (Table 8): HLM results for Trial 1

```
title1" trial1 results";
proc glimmix data=one_trail1 plots=residualpanel plots=boxplot maxopts=500 ;
            class donor segment group  id ;
            model score= group|day ;
            random int day /subject=donor;
            random int DAY/subject=segment(donor);
            lsmeans group/ pdiff diff plot=meanplot cl;
run;
```

Phase I: Trial 2 (Table 13): HLM results for Trial 2

```
title1" trial2 results";
proc glimmix data=one_trail2 plots=residualpanel plots=boxplot ;
            class donor segment group  id ;
            model score= group|day ;
            random int DAY/subject=segment(donor);
            random int / subject=id*segment(donor);
            lsmeans group/ pdiff diff plot=meanplot cl;
run;
```

Phase I: Trial 2 (Table 15): Testing group*week interaction found in Table 15

```
data weeks;
set weeks;
lgscore=log10(score);
run;
Title1 "observers vs expert scores comparison (Score is log transformed) ";
proc glimmix data=weeks plots=residualpanel plots=boxplot ;
where secotion="winter";
          class donor segment group  id week ;
          model lgscore= group|week ;
          random segment(donor) group*segment(donor);
          lsmeans group|week/ pdiff diff adjust=bon
          plot=meanplot(sliceby=group connect) cl;
run;
```

Phase II: Photograph Trial (Table 19): HLM results

```
proc glimmix data=one plots=residualpanel plots=boxplot maxopt=500;
       class id photo photo_type group arf ;
       model lgscore = group arf(group)/ ddfm=kr;
       random id;
       lsmeans group arf(group) / pdiff diff plot=meanplot cl;;
run;
```

Phase II: Photograph Trial (Table 20): HLM with experience and photo type included

```
proc glimmix data=one plots=residualpanel plots=boxplot maxopt=500;
      class id photo photo_type group arf ;
      model lgscore = group photo_type arf(group)/ ddfm=kr;
      random photo/subject=id type=ar(1);
      lsmeans photo_type group arf(group) / pdiff diff plot=meanplot cl;;
run;
```

VITA

Kelly Sauerwein was born April 12, 1982 in Downey, CA to Alana and Thomas Sauerwein. She graduated from Ramona Convent Secondary School in 2000 and moved to New Orleans, LA to attend Tulane University where she graduated Cum Laude in 2004 with a Bachelors of Arts degree in Psychology and History. Thinking she wanted to be a research psychologist, Kelly received her Masters of Arts in Psychology from the University of California, Davis in 2007 before discovering her love of Anthropology in 2008. She moved to Austin, Texas and attended Texas State University as a post-grad student in Anthropology, which is where she met Dr. Michelle Hamilton and was soon accepted to their Master's program in Anthropology in 2009. She received her Master's in Anthropology in 2011 from Texas State University and enrolled as a doctoral student at the University of Tennessee, Knoxville in 2011. While at UTK, she was both a graduate teaching assistant for the undergraduate introductory biological anthropology course and a graduate research assistant on various projects. She also participated as a lecturer and field leader for courses at the Forensic Anthropology Center and was involved in casework. Kelly received her Ph.D. in Anthropology under Dr. Dawnie Steadman in 2018.